

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C12N 15/12, C12Q 1/68, C07K 14/465, 14/47, A01K 67/027, C07K 16/18		A1	(11) International Publication Number: WO 96/39505 (43) International Publication Date: 12 December 1996 (12.12.96)
(21) International Application Number: PCT/GB96/01341 (22) International Filing Date: 5 June 1996 (05.06.96) (30) Priority Data: 9511439.3 6 June 1995 (06.06.95) GB (71) Applicant (for all designated States except US): ISIS INNOVATION LIMITED [GB/GB]; 2 South Parks Road, Oxford OX1 3UB (GB). (72) Inventors; and (75) Inventors/Applicants (for US only): GRIFFITHS, Richard [GB/GB]; 4 Upway Road, Oxford OX3 9QH (GB). TIWARI, Bela [CA/GB]; 4 Upway Road, Oxford OX3 9QH (GB). (74) Agent: PENNANT, Pyers; Stevens Hewlett & Perkins, 1 Serjeants' Inn, Fleet Street, London EC4Y 1LL (GB).		(81) Designated States: AU, IL, JP, NZ, US, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). Published <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>	
(54) Title: AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS			
(57) Abstract The invention provides for DNA sequences (introns and exons) encoding two genes located on the Z and W chromosomes of birds. These genes can be used to identify the sex of most birds (Class Aves). In addition, the genes control the sex of the birds, a genetic process which is initiated as an embryo. Nucleic acid fragments, proteins, polypeptides, antibodies and related products and their use in medicine and agriculture are provided. The invention may be used in sex identification or for controlling the sex of adults or the progeny of commercially important animals.			

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgyzstan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic	KR	Republic of Korea	SE	Sweden
CG	Congo	KZ	Kazakhstan	SG	Singapore
CH	Switzerland	LI	Liechtenstein	SI	Slovenia
CI	Côte d'Ivoire	LK	Sri Lanka	SK	Slovakia
CM	Cameroon	LR	Liberia	SN	Senegal
CN	China	LT	Lithuania	SZ	Swaziland
CS	Czechoslovakia	LU	Luxembourg	TD	Chad
CZ	Czech Republic	LV	Latvia	TG	Togo
DE	Germany	MC	Monaco	TJ	Tajikistan
DK	Denmark	MD	Republic of Moldova	TT	Trinidad and Tobago
EE	Estonia	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	UG	Uganda
FI	Finland	MN	Mongolia	US	United States of America
FR	France	MR	Mauritania	UZ	Uzbekistan
GA	Gabon			VN	Viet Nam

Avian CHD genes and their use in methods for sex identification in birds

5

Introduction

The present invention relates to proteins, polypeptides, nucleic acid fragments, antibodies and related products and to their use in medicine and agriculture, for instance in diagnosis and therapy. More particularly the invention relates to a gene or genes which can be used to ascertain the sex of avian adults, embryos, cells, and tissues. These genes also control the sex of birds starting with action in the embryos and so control the sex of the progeny of birds

Much of our understanding of sex determination comes from three, extensively studied, model systems. In two of these, the fruitfly *Drosophila melanogaster* and the nematode *Caenorhabditis elegans*, it is the ratio of X chromosomes to autosomes that initiates sexual differentiation (Hodgkin 1992). In the mouse a single gene, *SRY*, located on the Y chromosome provides the impetus for male development; a pattern that is thought to be conserved throughout the mammals (Koopman *et al.* 1991 Foster, *et al.* 1992).

At the genetical level these three species employ very different molecular mechanisms, not only to control sex determination itself but to accommodate the differing dosages of genes that result from the males possessing a single X and the female two X chromosomes. These basic differences are largely due to the independent evolution of the three mechanisms and strongly suggests that other means of sex determination will have evolved elsewhere in the animal kingdom.

One class in which little is known about sex determination is the birds. They exhibit female heterogamety which means that the female

has Z and W sex chromosomes and the male ZZ. This immediately suggests that sex determination in this class has an independent origin to that of their sister class, the mammals where it is the male that is heterogametic. Furthermore, it has been shown that whilst female
5 mammals inactivate one of their X chromosomes as a method of dosage compensation (Grant & Chapman 1988), this does not seem to be a device employed by birds (Baverstock *et al.* 1982).

However, similarities do exist between the birds and mammals. The W chromosome, like the Y chromosome is usually smaller
10 than its partner, and is also characteristically heterochromatic in appearance (Christidis 1990). The main exceptions to this rule are found in the 'primitive' representatives of both classes: the monotremes and the ratites where the morphological differences between the sex chromosomes are poorly defined (Graves 1987, Tagaki *et al.* 1972).

15 The heterochromatization of the W and Y results from the replacement of functional genetic loci with 'junk DNA' sequences. This process is thought to be a consequence of a suppression of recombination that has arisen to ensure that genes vital to the development of the heterogametic sex remain linked on the Y or W chromosome
20 (Charlesworth 1991). As a result only a few genes such as *Ube1y* (Kay *et al.* 1991, Mitchell *et al.* 1991), *Zfy* (Page *et al.* 1987) and *SRY* itself remain on the mammalian Y chromosome. A similar situation is thought to prevail on the avian W chromosome where the presence of any functional genes has yet to be demonstrated, although it does possess vast arrays of
25 repetitive elements (Griffiths & Holland 1990, Tone *et al.* 1982).

A further similarity in sex determination in birds and mammals is that the development of the male phenotype appears crucially dependent on the appearance of the testis. The female phenotype is the result of the 'default pathway'. For mammals this was first demonstrated
30 by Jost (1947) who grafted an embryonic testis into genetically female

rabbit embryos prior to sex determination. This was sufficient to allow the development of functional males. The same experiment has been carried out on chick embryos with comparable results (Stoll *et al.* 1978).

Once the testis has formed, the process of masculinization is adopted by the testicular hormones. The genetical switch that initiates testis determination is known to be *SRY* in mammals (Koopman *et al.* 1991). In birds, there appears to be no *SRY* homologue on the W chromosome (Griffiths 1991), although this is unsurprising given the separate evolution of sex determination in the two classes.

The only other pertinent evidence on the genetics of avian sex determination come from reports of chickens with abnormal chromosome complements. Table 1 shows data from Crew (1954) and McCarrey and Abbott (1979) on the phenotypes of the aneuploids so far described. These results suggest that the presence of the W chromosome in the aneuploid AA ZZW and the polyploid AAA ZZW has not acted as a dominant determinant of the female phenotype. This may mean that sex in birds may be determined more by the autosome to Z ratio, as in *Drosophila* and *C. elegans*. However, a ZO aneuploid which could confirm this hypothesis has yet to be described.

It must also be born in mind that XXY kangaroos, where *SRY* is thought to be the key male determining switch, exhibit both male and female characteristics (Graves 1987). This suggests that the limited aneuploid data that is available for birds should be interpreted with some caution.

To conclude, the genetic mechanism that controls sex determination in birds has not yet been elucidated. Here we suggest that a gene we have termed *CHD-W* (Chromodomain-*Helicase*-DNA binding on the W chromosome) alone or acting in conjunction with a closely related gene *CHD-1A* (Chromodomain-*Helicase*-DNA binding 1 Avian) initiates female development in birds.

The Invention

It is believed that all birds such as chickens and other species of commercial significance, will have two or more genes of the CHD type which will have a nucleotide sequence similar to the nucleotide sequences shown in Fig. 5, Fig. 7 and Fig. 8 and that the gene products will be proteins which are crucial to the determination of the sex of the organism. One of these genes will be located on the W chromosome and the other on an autosome or Z chromosome.

It will be understood that the exact sequence of the two genes will vary between species and between individuals of the same species at least at the nucleotide level and often also at the protein level. Complete or partial sequences of the chicken genes are shown in Fig. 5, Fig. 7 and Fig. 8. The gene or protein which contains sequence corresponding to those in Fig. 5, Fig. 7 and Fig. 8 will hereafter be referred to as an CHD-gene and proteins and fragments thereof, polypeptides, nucleic acids and fragments thereof and oligonucleotides containing part of a CHD gene will hereafter be referred to as CHD-proteins, CHD-nucleic acids and so on.

The present invention therefore provides a CHD-protein or a fragment thereof or polypeptide comprising a CHD-gene or a part thereof, subject to the proviso below.

The present invention also provides a protein or a fragment thereof or a polypeptide containing a mimetope of an epitope of a CHD-protein or fragment thereof of polypeptide containing a CHD-gene or a part thereof, subject to the proviso below. Such proteins, fragments and polypeptides are hereafter referred to as CHD-mimetope proteins or fragments thereof and CHD-mimetope polypeptides.

The present invention also provides a CHD-nucleic acid or a fragment thereof or oligonucleotide comprising a CHD-gene, or a part thereof subject to the proviso below.

In a particular aspect the present invention provides a single or double stranded nucleic acid comprising the CHD-gene of a bird or a part thereof of at least 17 contiguous nucleotide bases or base pairs, or a single or double stranded nucleic acid hybridizable with the CHD-gene of a bird, or part thereof of at least 17 contiguous nucleotide bases or base pairs, subject to the proviso below.

The invention further provides a nucleic acid or fragment thereof or an oligonucleotide encoding a CHD-protein or fragment thereof or a polypeptide comprising a CHD-gene or a part thereof or a CHD-mimetope protein or a fragment thereof or CHD-mimetope polypeptide, subject to the following proviso. These nucleic acids, fragments and oligonucleotides may have sequences differing from the sequences of CHD-nucleic acids, fragments and oligonucleotides due to alternative codon usage and/or encoding alternative amino acids sequences or mimetopes.

The present invention does not, however extend to any known protein or fragment thereof or polypeptide or nucleic acid or fragment thereof or oligonucleotide containing a CHD-gene related sequence such as the *Saccharomyces cerevisiae* SNF2/SWI2 gene, *Drosophila polycomb* and HP1 genes described below, insofar as that protein or fragment, polypeptide, nucleic acid or fragment or oligonucleotide is known *per se*.

The amino acid sequence of the CHD-gene has similarities to the chromobox and *Helicase* motifs of a number of discovered genes known to be involved in the remodelling of chromatin. This suggests that the CHD-protein of the present invention may have a regulatory function involving chromatin remodelling. However, none of these genes contain the chromobox and the *Helicase* of the CHD-gene which are conserved in conjunction, at least in the chicken, great tit, mouse and yeast but are not conserved in conjunction in the sequences of chromatin remodelling

proteins not associated with sex determination at least at the stage of testis formation in birds. A gene that produces a protein having chromatin remodelling capacity but lacking these characteristic motifs is therefore outside the scope of the present invention.

5 In addition there are certain residues in the amino acid sequence of the chromobox and those residues immediately downstream thereof, of the CHD-gene which are also conserved at least between those found in the chicken, great tit, mouse and yeast but are not conserved in the sequences of chromatin remodelling proteins not associated with sex
10 determination at least at the stage of testis formation in birds. Any one of these conserved residues is therefore considered characteristic of the CHD-gene proteins of the present invention. The characteristics of a CHD-chromobox will give a more complete and comprehensive description of the CHD-chromobox which can also be considered characteristic of the
15 CHD-gene proteins of the present invention. A protein having chromatin remodelling capacity and a helicase motif but originating from a gene that lacks all or most of these characteristic amino acid residues in the chromobox motif is therefore outside the scope of the present invention.

The characteristic amino acids residues are shown in the
20 alignment in Fig. 11, which is described in more detail below. When aligned with the illustrated sequences as shown, these residues fall at positions, 11, 12, 20, 27, 34 inside the chromobox and 3, 6, 8, 12-15, 16 immediately downstream.

The nucleotide base sequence of the CHD-gene includes
25 bases which encode the chromobox and *Helicase* motifs of chromatin remodelling proteins as described above. However, the base sequence of the CHD-nucleic acids of the gene will include codons specifying both or either chromobox and *Helicase* motifs and the former will have codons specifying one or more of the characteristic amino acid residues described
30 above and/or will be hybridizable with a sequence that controls the sex

determination of birds under conditions which substantially prevent hybridization to other sequences in birds that do not have these characteristics.

Preferably the CHD-nucleic acids of the invention encode a
5 chromobox and a helicase and one or more, preferably all, of the characteristic chromobox amino acid residues and meet the above hybridization requirements.

Fragments of CHD-nucleic acids according to the present invention will likewise contain codons specifying the chromobox and
10 helicase motifs or including at least part of either of these motifs or CHD-gene adjacent to the codons encoding these features and/or will be hybridizable with a sequence that controls the sex determination of birds under conditions which substantially prevent hybridization to other sequences in birds that do not have these characteristics.

15 Oligonucleotides containing the CHD-gene or a part thereof according to the present invention may contain codons specifying the chromobox or helicase motifs or including at least part of these motifs or CHD-gene but this is not essential. However all such oligonucleotides of the invention must be capable of hybridizing with a sequence or sequences
20 that control the sex determination of birds or a gene intron, preferably under conditions which substantially prevent hybridization with any sequence not associated with sex determining sequence.

A sex determining sequence referred to herein is a sequence which contains the CHD-gene and which encodes a factor which when
25 expressed at the appropriate stage and level during embryo development may result in testis formation and subsequent growth of the embryo as a male. It may alternatively refer to a sequence which encodes a factor which when expressed at the appropriate stage and level during embryo development prevents testis formation and results in the subsequent
30 growth of the embryo as a female.

The hybridization conditions referred to above which prevent unwanted hybridization with sequences not associated with the sex determining gene will depend to some extent on the length of the nucleic acid, fragment or oligonucleotide of the invention tested. Thus for instance
5 lower stringency will be sufficient to secure hybridization to sequences associated with the sex determining gene whilst preventing unwanted hybridization when the nucleic acid or fragments several thousand nucleotide base pairs in length than for a fragment of only a few hundreds of bases or an oligonucleotide of from 17 bases up to a few tens or
10 hundreds of bases. With the smallest oligonucleotides and fragments of the invention hybridization conditions will be such that only complete complementarity between the oligonucleotide and or fragment and the sequences associated with the sex determining gene will result in hybridization.

15 Preferred nucleic acids and fragments of the invention will only hybridize selectively to the sequences associated with the sex determining gene or genes under conditions requiring at least 80%, for instance 85, 90 or even 95% more preferably 99% complementarity. Yet more preferred nucleic acids and fragments of the invention are those
20 having a sequence corresponding exactly to that of those illustrated in Fig. 5, Fig. 7 and Fig. 8 although the nucleotide sequences may be longer or shorter than those illustrated and or may contain normally intronic sequences associated with these sequences

The invention particularly provides an oligonucleotide,
25 polypeptide, nucleic acid or protein comprising the entire sequence of the CHD-gene of a bird and more preferably comprising the entire amino acid or nucleotide sequence of the chicken as set out in any one of Figs. 1, 3, 5, 7, 8, 9, 10, 11.

The nucleic acids hybridizable with the CHD-gene of a bird
30 are preferably hybridizable under moderate, or more preferably, high

stringency conditions as defined below:

Moderate stringency:

Buffer: 2 x SSC
5 Temp: 50°C
Annealing period: 6-8hrs

High stringency:

Buffer: 1 x SSC
10 Temp: 65°C
annealing period: 6-8hrs

Moderate stringency as defined above corresponds with about 75% homology. High stringency as defined above corresponds with about 90% homology. 1 x SSC is 0.15 M sodium chloride, 0.015 M sodium citrate, pH 7.0.

Preferably the portion of the nucleic acid corresponding to or hybridizable with the CHD-gene is at least 20, more preferably at least 30, 40 or 60 and most preferably 100 or more nucleotide bases in length.

20 The nucleotide strands of the invention may be single or double stranded DNA or RNA. DNA's of the invention may comprise coding and/or non-coding sequences and/or transcriptional and or translational start and/or stop signals and/or regulatory, signal and/or control sequences such as promoters, enhancers and/or polyadenylation sites, endonuclease restriction sites and/or splice donor and/or acceptor, in addition to the CHD-gene sequence. Included within the DNA's of the invention are genomic DNA's and complementary DNA's (cDNA's) including functional genes or at least an exon containing the CHD-gene. They may also contain non-coding sequences such as one or more
30 introns. Single stranded DNA may be the transcribed strand or the non-

transcribed (complementary) strand. The nucleic acids may be present in a vector, for instance a cloning or expression vector, such as a plasmid or cosmid or a viral genomic nucleic acid. RNA's of the invention include unprocessed and processed transcripts of DNA, messenger RNA (mRNA)
5 containing the CHD-gene and anti-sense RNA containing a sequence complementary to the CHD-gene.

Nucleic acids of the present invention are particularly useful as primers for polymerase chain reactions (PCRs) conducted to ascertain the sex of a bird as defined below. They may also be used to express
10 proteins or fragments or polypeptides corresponding to the whole or a part of a CHD-protein (whether or not containing a CHD-gene) or as probes in hybridization experiments. As used herein the term "fragments" used in connection with proteins is intended to refer to both chemically produced and recombinant portions of proteins.

15 The CHD-proteins and fragments thereof and polypeptides containing the CHD-gene or a part thereof and CHD-mimotope proteins and fragments thereof and CHD-mimeotope polypeptides of the invention are useful in immunodiagnostic testing and for raising antibodies such as monoclonal antibodies for such uses. Antibodies against such proteins
20 and fragments and polypeptides as well as fragments of such antibodies (which antibody fragments include at least one antigen binding site) including chemically derived and recombinant fragments of such antibodies, and cells, such as eukaryotic cells, for instance hybridomas and prokaryotic recombinant cells capable of expressing and, preferably
25 secreting antibodies or fragments thereof against such proteins or fragments, also form part of the present invention.

The nucleic acids of the invention may be obtained by conventional means such as by the recovery from organisms using PCR technology or hybridization probes, by *de novo* synthesis or a combination
30 thereof, by cloning the CHD-nucleic acids described below or a fragment

- 11 -

thereof or by other techniques well known in the art of recombinant DNA technology.

Proteins and fragments thereof and polypeptides of the invention may be recovered from cells of organisms expressing a CHD-gene or generated by expression of a CHD-gene or coding sequence
5 contained in a nucleic acid of the present invention in an appropriate expression system and host, or obtained by *de novo* synthesis or a combination thereof, by techniques well known in the art of recombinant DNA technology. The proteins, fragments thereof and polypeptides of the
10 invention will contain naturally occurring L-a-amino acids and may also contain one or more non-naturally occurring a-amino acids having the D- or L- configuration

Antibodies may be obtained by immunization of a suitable host animal and recovery of the antibodies, by culture of antibody
15 producing cells obtained from suitably immunized host animals or by *in vitro* stimulation of B-cells with a suitable CHD-protein, fragment or polypeptide or CHD-mimotope, protein, fragment or polypeptide and culture of the cells. Such cells may be immortalized as necessary for instance by fusion with myeloma cells. Antibody fragments may be
20 obtained by well known chemical and biotechnological methods.

All these techniques are well known to practitioners of the arts of biotechnology. Reference may particularly be made to the well known text book "Molecular cloning: A laboratory manual" 2nd Edition (Eds Sambrook, J., Fritsch, E.F. and Maniatis, T.), (Cold Spring Harbour
25 Laboratory, New York, 1989), hereafter referred to as "Maniatis".

The invention further provides the use of a nucleic acid, protein, polypeptide, antibody, or antibody producing cell as hereinbefore defined including the SNF2/SWI2, polycomb and HP1 or other chromobox or helicase containing protein for ascertaining the sex of a cell or organism
30 of a bird or for isolating nucleic acids useful in ascertaining the sex of a bird

and for instituting single sex breeding programmes.

Knowledge of the chicken or great tit sex determining gene or genes can be used to isolate the equivalent gene or genes from other birds. Once isolated from a particular species, this gene or genes and its
5 sequence can typically be used in two types of application:

1. The construction of sequence based sexing tests which can be applied to embryos, tissues and other biological materials containing nucleic acids.
2. The genetic modification of the germ line of birds to create
10 breeding systems that produce offspring statistically biased towards one sex or of one sex only (single sex breeding systems).

A particularly preferred technique for ascertaining the sex of a bird in accordance with the invention involves the use of an oligonucleotides as primers in a PCR, for instance as follows:

15 A cell or cells or remains thereof are obtained, for instance by surgical removal from an embryo or from the quill of a feather, and the DNA is released by a crude lysis procedure for instance using a detergent or by heating. Primer oligonucleotides of the invention are used to initiate a conventional PCR in order to amplify W chromosome linked CHD-related
20 DNA from the cells. The products of the PCR are analysed by agarose gel electrophoresis and detected using labelled probes or by visual inspection. The presence of amplified CHD-W DNA indicates the presence of a CHD-W gene in the cells and thus, in birds, that the cell(s) were female. An example of a similar technique has been carried out by Griffiths & Tiwari
25 (1995) on the Spix's Macaw (*Cyanopsitta spixii*). This is the world rarest bird (Guinness Book of Records) and DNA obtained from a moulted feather was sufficient to allow nested PCR amplification with CHD primers to show the bird was a male.

This technique may be applied for instance to identify the sex
30 of embryos or adults for subsequent breeding programs in other bird

- 13 -

species, or to control the sex of the progeny of breeding stock for commercial exploitation (by selection of the breeding stock or by slaughter or termination of animals of undesired sex).

5 The oligonucleotide primers for ascertaining or controlling sex in one species may also be used to ascertain or control sex in another species since hybridization of the primers to the CHD-gene of the other species will still serve to amplify the species-specific sequences.

Techniques for conducting such determinations are well known in the art of recombinant DNA technology.

10 In another aspect the present invention provides a process for isolating a W-chromosome specific sequence associated with the *CHD-W* gene of a bird which comprises probing a genomic library from a female of the species preferably of W chromosome sequences, for instance of lambda phage, cosmid or YAC library or cDNA library constructed from a
15 tissue expressing the gene, with a probe comprising a nucleic acid, fragment or oligonucleotide of the invention as hereinbefore defined and a detectable label under high or moderate stringency.

Using the newly isolated subclone, Southern blots are performed on male and female DNA of the species of interest at high
20 stringency to confirm that the correct clone has been isolated. The CHD-gene probe should give a female specific signal (other male/female shared bands may also be present at lesser intensities). The subclone is sequenced using standard methods and primers suitable for PCR chosen from the sequence so identified.

25 Alternatively, other approaches to cloning the sequences related to the sex determining gene could be used such as PCR methods using "degenerate" oligonucleotides. (For methods in PCR see, for example, "PCR Protocols - a Guide to Methods and Application"; edited by M.A. Innis, D.H. Gelfand, J.J. Sninsky, T.J. White; published by Academic
30 Press, Inc.).

Preferably the probe is CHD-1A or *CHD-W* or a fragment thereof or a nucleic acid or fragment or oligonucleotide having a sequence exactly as set out in Fig. 5, Fig. 7 or Fig. 8 for the chicken. Techniques for forming a genomic or cDNA library and for probing and detecting the detectable label and isolating the nucleic acid identified by the probe are well known in the art of biotechnology and recombinant DNA manipulation. The process may be conducted for instance using a probe having the chicken sequence such as the *CHD-W* sequence to identify and isolate the corresponding sequence from another bird such as Turkey. The thus-identified sequence can then be used to generate primers for PCR which in turn can be used to ascertain the sex of an individual or of cells, tissues, embryos or ovaries of the bird. This technique has been used by obtaining DNA from the Chicken and Hyacinth Macaw (*Anodorhynchus hyacinthinus*) to design primers for the Spix's Macaw (Griffiths & Tiwari 1995). This will permit experiments to ascertain sex to be conducted and controlled sex breeding of the bird as described below.

In addition, the nucleotide sequence of the CHD-genes are sufficiently conserved so that CHD primers can be designed that will allow PCR in a range of bird species. The primers P1, P2 and P3 shown in Figure 14 will allow *CHD-W* and *CHD-1A* amplification in a range of birds that allows sex to be identified.

The isolated nucleic acid, fragment or oligonucleotide may thereafter be amplified, cloned or sub-cloned as necessary. The invention further provides a process for detecting the sex of an individual bird or of cells, tissues, embryos, foetuses or ovaries or a bird, comprising conducting a polymerase chain reaction using DNA from the individual, cell, tissue, embryo or ovary as template and a nucleic acid, fragment or oligonucleotide of the invention as primer. Preferably the nucleic acid, fragment or oligonucleotide of the invention used as primer is *CHD-W* or *CHD-1A* or a part thereof and has a sequence corresponding exactly to the

chicken sequence in Fig. 5, Fig. 7 or Fig. 8 or a part thereof or is a nucleic acid, fragment or oligonucleotide which is a W-chromosome specific sequence associated with the sex determining gene or genes of a bird of the same species as the individual cell, tissue, embryo, foetus or ovary whose sex is to be ascertained. The W-chromosome specific sequence associated with the sex determining gene or genes of the bird involved may itself have been obtained by the process of isolation and amplification or cloning described above. It can also be obtained by deduction from the sequence in Fig. 5, Fig. 7 or Fig. 8 or a sequence from another bird or animal.

The identification of the sex determining gene or genes according to the present invention raises the possibility of controlling the sex of progeny of commercially important animals such as chickens, turkeys and other avians. This will be valuable in many aspects of animal breeding and husbandry such as where one sex has more desirable characteristics, for instance only female progeny are desired for egg-laying breeds of chicken. The economic advantages of single sex breeding programmes and strategies for instituting these are described for instance in "Exploiting New Technologies in Animal Breeding; Genetic Developments", (Eds. Smith, C., King, J.Q.B. and McKay, J.C.), (Oxford University Press, Oxford, 1986).

The nucleic acids making up all or part of the sex determining gene, from the same or different animal species, can be introduced into any early embryo through established transgenic technology. This latter includes microinjection of DNA into pronuclei or nuclei of early embryos, the use of retroviral vectors with either early embryos or embryonic stem cells, or any transformation technique, (including microinjection, electroporation or carrier techniques) into embryonic stem cells or other cells able to give rise to functional germ cells. These procedures will allow the derivation of individual transgenic animals (founder transgenics) or

- 16 -

chimeric animals composed in part of cells carrying the introduced DNA. Where the functional germ cells of the founder transgenic or chimeric animal carry the introduced DNA it will be possible to obtain transmission of the introduced DNA to offspring and to generate lines or strains of animals carrying these DNA sequences.

The nucleic acids making up part or all of the coding sequence of the sex determining gene, or derivatives of it, may be introduced in combination with its own regulatory sequences (promoter/enhancers etc.) or regulatory sequences from another gene, the whole making the "construct", to give expression from the construct at an appropriate developmental stage and tissue location critical to sex determination in the bird species under consideration. For example, in the chicken this would be between 6 and 7 days post lay.

15

Materials and Methods

Isolation of pGT-W, pGT1.7 and pGT8 *Great Tit* clones

A great tit (*Parus major*) library was constructed from genomic DNA, partially restricted with *MboI*, and the IFixII vector (Stratagene). The library was screened at high stringency with the 724bp probe (GT-W) cloned from a W chromosome specific polymerase chain reaction (PCR) product derived from the great tit (Griffiths & Tiwari 1993). Positive plaques were subject to two rounds of purification. Clone IGT2 contained an insert of 9.6kb that hybridized strongly to the probe sequence. The insert was subcloned as two *EcoRI* fragments of 1.7kb (pGT1.7) and 8kb (pGT8) into *EcoRI* cut pT7/T3 (Pharmacia).

Isolation of CHD genes from the chicken

Two chicken cDNA libraries were screened. The first was a mixed sex chick stage 10-12 cDNA library in lZapII which had been reamplified on 2

- 17 -

occasions This library was provided by Dr I. J. Mason. The second library was constructed from mixed sex, 10 day chick mRNA. Total RNA was extracted using a guanidine thiocyanate based technique (Koopman 1993) and mRNA isolated using a Promega PolyATtract system 1000. A lZapII library was constructed using a Stratagene ZAP-cDNA synthesis kit. Plaques (2×10^5) from the stage 10-12 day library were screened at moderate stringency with a subcloned 433bp HindIII/SacI fragment from pGT8 that contained the 123bp region with identity to the mouse *CHD-1* gene (Delmas *et al.* 1993). A similar number of plaques from both libraries were screened with bases 428-4428 of *CHD-1A* (see Fig. 5). The 10 day library was also screened with bases 4059-5303 of *CHD-1A* (see Fig. 5). Positive plaques were purified prior to the excision of pBluescript plasmids and cloned inserts insert from lZapII using techniques recommended by Stratagene.

15

Sequencing

All sequencing was carried out using the T7 DNA polymerase/7-deaza-dGTP chain termination sequencing kit from USB. All sequencing unless otherwise specified was carried out in both directions either by subcloning or through exonuclease III deletion with the Promega Erase-a-Base system.

20

Southern Blot Analysis and Hybridization

Genomic DNA was extracted from blood (Griffiths & Holland 1990), digested with the appropriate restriction enzyme and Southern blotted onto Zeta-Probe GT under neutral conditions as described by the manufacturer (Bio-Rad). Prehybridizations and hybridizations were carried out in 0.25M Na_2HPO_4 /5% SDS at either 65°C (high stringency) or 62°C (moderate stringency). Subsequent washes were carried out for a total of 1 hour in three changes of either 0.5 x SSC (75mM NaCl/7.5mM sodium

30

citrate (pH7.5))/0.1%SDS at 65°C (high stringency) or 1 x SSC/0.1%SDS at 45°C (low stringency).

**Sex identification with PCR on dried and limited DNA in a
5 Spix's Macaw**

Stratagene provided a genomic Hyacinth Macaw Lambda
FixII Library (Cat. No. 946402). Plaques were screened at moderate
stringency with a 1.3Kb Chicken *CHD-W* subclone (spans 2670-4003
nucleotides in the related Mouse *CHD1* gene (Delmas *et al.*, 1993)). A
10 *CHD-W* genomic fragment was isolated and aligned to the chicken and
mouse homologues to allow the design and construction of 3 primers (5' to
3') P3 AGATATTCCGGATCTGATAGTGA,
P2 TCTGCATCGCTAAATCCTTT and
P1 ATATTCTGGATCTGATAGTGA(C/T)TC.

15 DNA from the wild Spix's Macaw was extracted (Thomas &
Pääbo 1993) from 1cm portions of the tips of 3 moulted flight feathers
collected in 1994 and 1995. The negative extraction control was taken
through an identical procedure. 1.5% of these extraction products or 50ng
of genomic DNA from the reference samples were subject to semi-nested
20 PCR. Primary amplification consisted of 20 cycles with primers P3 and P2;
1% of the primary PCR product was subject to 30 cycles of amplification
with P2 and P1. Samples were denatured for 1.5 min at 95°C then cycled
between 57°C/30 sec, 72°C/15 sec and 94°C/30 sec with a 5 min final
extension. Products were precipitated, cut with *DdeI*, reprecipitated and
25 electrophoresed through visigel separation matrix (Stratagene). The
accuracy of the test was confirmed using DNA from Spix's and Hyacinth
Macaws of known sex (n=5 p=0.03). Uncut secondary PCR product from
the wild bird was isolated (Dretzen *et al.* 1981), cloned using the
Stratagene pCR-Script SK(+) kit and sequenced to confirm that the product
30 had originated from a Spix's Macaw

Sex identification with PCR in a variety of birds

DNA was isolated from blood taken from Chicken (5 individuals used), Marsh Harrier (28; *Circus aeruginosus*) and Kestrel (18 *Falco tinninculus*) all sexed by adult plumage, Bee-eater (4; *Merops*
5 *apiaster*, plumage/behaviour), Boobook Owl (2; *Ninox novaesiae*), White-faced Owl (2; *Ptilopsis leucotis*) Burrowing Owl (2; *Speotyto cumcularia*), Eurasian Eagle Owl (2; *Bubo bubo*), Long-eared Owl (2; *Asio otus*), Tawny Owl (3; *Strix aluco*, adult size), Starling (5; *Sturnus vulgaris*; Beak colour) and African Marsh Warbler (5; *Acrocephalus baeticatus*; reproductive
10 behaviour). DNA from a variety of parrots sexed by laparotomy was also used: Blue Fronted Amazon (3; *Amazona a aestiva*), Orange Winged Amazon (5; *Amazona amazonica*), Red Lored Amazon (3; *Amazona autumnalis*), Yellow Crowned Amazon (2; *Amazona o ochrocephala*), Tucamen Amazon (2; *Amazona tucamana*), Blue and Gold Macaw (6; *Ara*
15 *ararauna*), Citron Crested Cockatoo (2; *Cacatua sulphurea citronocristate*), Lesser patagonian (2; *Cyanolisous patagonus*), Blue Headed Pionus (1; *Pionus menstruus*), Plum Headed Parakeet (4; *Psittacula cyanocephala*), African Grey Parrot (12; *Psittacus erithacus*), Blue Throated Conure (2; *Pyrrhura cruentata*), Senegal Parrot (3; *Seneglus poicephalus*).

20 All the birds listed above were sexed from DNA using exactly the same PCR reaction. PCR reaction volumes of 20 μ l were made up of Promega Taq buffer (1x is 50mM KCl, 10mM Tris.HCl, 1.5mM MgCl₂, 0.1% Triton X-100), 200 μ M of each dNTP, P2 (5'-TCTGCATCGCTAAATCCTTT) and P3 (5'- AGATATTCCGGATCTGATA) primers (approx 1 μ M), 50-200ng
25 of genomic DNA and 0.15 units of Taq polymerase. The thermal treatment was 94°C/1.5mins followed by 30 cycles of 55 or 56°C/15sec, 72°C/15sec, and 94°C/30sec with a finish of 56°C/1min and 72°C/5min. *HaeIII* (5 units; Promega) was used to cut 8 μ l of PCR product in 1x Promega restriction enzyme buffer 3 and 50ng/ μ l bovine serum albumin (Sigma) in a total
30 volume of 10 μ l. The digests and uncut PCR product were precipitated

before being electrophoresed in a visigel (Stratagene) with ethidium bromide (40ng/ml) at 3.5V/cm.

Results

5 The plasmid pGT-W contains a 724bp insert that hybridizes to a 4.9kb fragment only in the female great tit. Its DNA sequence was determined (Fig.1) and contains a 457bp open reading frame. A search of the EMBL DNA and protein sequence database found no significant matches. The sequence does contain a simple sequence consisting of a
10 22bp run of thymidines.

 The pGT-W insert was used to probe Southern blots, at low stringency, of PvuII restricted genomic DNA of male and female great tit, starling, jackdaw (*Corvus monedula*), pied wagtail (*Motacilla alba*) and a species of new world flycatcher. These are species that cover the
15 extremes of the passeriforme order according to the recent phylogeny of Sibley *et al.* (1988). In all but the jackdaw convincing hybridization to a single female specific fragment could be observed. In all species, hybridization to one or more non-sex specific fragments was also shown. A similar experiment was carried out with a non-passerine, the bee-eater
20 (*Merops apiaster*), and this too resulted in faint hybridization to a female specific fragment and two, somewhat stronger bands, in both sexes.

 In order to further investigate the nature of the pGT-W insert we attempted to clone a larger fragment of genomic DNA which incorporated this motif. From around 1.5×10^5 plaques from a great tit
25 genomic library, two positives were obtained. After purification one of these gave superior hybridization and was investigated further. The 9.7kb insert was subcloned as pGT1.7 and pGT8 containing 1.7kb and 8 kb respectively. The pGT1.7 was sequenced in its entirety and approximately 2.8kb of the sequence of pGT8 was determined. Both were sequenced in
30 a single direction. A 723bp region, starting 133bp from the 5' end of pGT8

had a sequence that corresponded exactly to the pGT-W insert (Fig. 2).

The sequences derived from these subclones were used to search the EMBL database using the FASTA algorithms (GCG, Wisconsin package vers 7.3). A region of 123bp, starting 994bp from the 5' end of pGT8, showed a 79% nucleotide sequence identity to bases 3855-3977 of the mouse *CHD-1* gene (Fig. 3; Delmas *et al.* 1993). This corresponds to an 88% identity at the amino acid level.

Southern blots of *PvuII* digests of genomic DNA from male and female chicken and lesser black-backed gull (*Larus fuscus*) were probed at low stringency with a 433bp *SacI/HindIII* fragment of pGT8 that included the 123bp region with *CHD-1* identity (Fig. 4). Figure 12 shows that in the chicken hybridization was with a fragment of 3.1kb in the female only and with fragments of 1.5 and 6.0kb in both sexes. In the gull hybridization is similarly with a female specific fragment of 4.0kb a fragment of 3.0kb in males and females.

Delmas *et al.*, (1993) have already demonstrated the universal occurrence of the *CHD-1* in the mammals. The evidence this blot provides, which features species representing both the major divisions of the birds, suggests that a minimum of two types of CHD gene exist in this Class. The first we termed *CHD-W* to denote its W linkage. The 123bp region from the great tit would appear to be a short exon from this gene. The second hypothetical gene is closely related to *CHD-W* and we have it termed *CHD-1A*, where the A denotes its avian nature. This gene is either Z or autosomally linked as it occurs in both sexes.

Isolation of *CHD-1A*

The *SacI/HindIII* great tit probe was used at low stringency to screen a lZap II cDNA library from stage 10-12 (33-49hrs after the appearance of the primitive streak) chicken embryos. A plating of 2×10^5 plaques yielded a panel of 25 positive clones, 19 of these continued to

hybridize intensely after purification. From three clones Z4, Z6 and Z11 a composite 6608 nucleotide sequence (Fig. 5) was determined using the strategy illustrated in Fig. 6.

The insert from the Z6 clone (bases 418-4426; Fig. 5) and a *Bgl*III (AGATCT) fragment of the Z4 clone (bases 4059-5303; Fig. 5) were used separately to screen a similar number of plaques from a second cDNA library constructed from 10 day old chicken embryos. This screening identified a total of 45 positives of which 16 were found to have sequence identity with the composite sequence derived from the first library. Two additional clones contained a closely related sequence that is dealt with below.

A proportion of the clones from both libraries show variation from the sequence given in Fig. 5 in one respect. Clones Z1, Z13, Z17, Z20 and Z23 are identical to the composite sequence 5' to base 4327 from there they terminate in an additional 37 to 163 bases of a new sequence that is identical in all five. Two clones from the second library CC43 and CC56 have 22 or 254bp of the same sequence at their 5' ends.

Downstream of this motif both clones regained homology with the composite sequence at base 4328 and show no further deviation from the original sequence. From these seven clones a composite 264bp sequence can be derived and this is illustrated in Fig. 7. None of the seven clones contain the whole of this sequence. Moreover, none of the ten clones that span the 4327/4328 insertion point contain any of this additional region. If inserted at this position, the motif has an in frame, open reading frame spanning its entire length. The motif is extremely adenosine rich and this makes the amino acid lysine extremely common in the putative translation (see Fig. 7). There are no splice donor or acceptor sites within the motif suggesting it is a final rather than an intermediary product of splicing.

Hybridization of a probe running from 2534 to 4428bp of the sequence chicken sequence to a blot of *Pvu*II cut, male and female

chicken genomic DNA shows that hybridization occurs to fragments that are both W and autosomally or Z chromosomally located. The level of hybridization is significantly stronger to the fragments common to both sexes suggesting that the probe represents the *CHD-1A* gene.

5 *CHD-1A* is very closely related to the mouse *CHD-1* gene being 79.8% identical in a 5152nt overlap. At the amino acid level the identity is raised to 90% over 1750 residues. We do have an additional 1202bp of the 3' untranslated region but have not encountered a clone with an AATAAA termination signal or a 3' homopolymeric T tail. Both mouse
10 and chicken sequences contain a stop codon in the same relative positions and sequence similarity is insignificant after this point. The published mouse sequence does not contain the additional 264bp motif described above.

 The database search also identified an unpublished chicken
15 derived sequence tagged as a delta crystallin binding protein (*DCBP*), with even greater identity than the mouse *CHD-1* gene: 99% over 2293 bp and 94% over 571 amino acid residues. The *DCBP* sequence is of 2292bp which extends over nucleotides 1922 to 4214 of *CHD-1A* (Fig. 5). Despite the high nucleotide sequence identity the region of amino acid similarity
20 does not extend the full length of the *DCBP*. This is due to apparent deletions in the *DCBP* clone that provides an initiation methionine codon (257nt *DCBP*) and a stop codon (1939nt *DCBP*). The extremely high sequence identity, the fact that identity is maintained after the apparent stop in the *DCBP* sequence, that none of the 41 *CHD*-related clones we
25 found have exact sequence identity and that only small sequencing mistakes would be required to introduce false stop and start codons suggests that the *DCBP* sequence is *CHD-1A* but has been sequenced slightly inaccurately. Further evidence is required to confirm this.

 The database search with the whole *CHD-1A* gene also
30 revealed significant identity to a previously unidentified portion of a 15 kb

region of *S. cerevisiae* chromosome V. This region comprises an open reading frame of 4.4kb which lies between the RAD4 (Gietz & Prakash 1988) and the poly-A binding protein (Sachs *et al.* 1986) gene coding regions. In an overlap of 1538 amino acids, the whole of the yeast open reading frame, there is an identity of 37.7% and a similarity of 59% (Fig. 10). The degree of conservation this similarity implies suggests the yeast sequence encodes a homologue of *CHD-1A* that we shall term *CHD-1Y* for the sake of discussion.

Delmas *et al.*, (1993) identified four motifs in *CHD-1* with possible functional significance. *CHD-1A* retains such close homology to *CHD-1* that these regions are virtually unchanged and are likely to perform similar functions as they do in the mouse.

The first motif is a chromodomain (Paro & Hogness 1991) which falls between residues 274 and 311 (Fig. 9). Figure 11 compares the amino sequence of this region to that of eight others identified through a search of the EMBL database. The sequences fall into three categories. The first comprises the domain from *CHD-1*, *CHD-1A* and *CHD-1Y*. The second and third chromobox groups have been previously identified by (Pearce *et al.* 1992). The HP1 class comprises the *Drosophila* (James & Elgin 1986) and human (Saunders *et al.* 1993) HP1 genes and two murine modifier (*Mod*) genes (Singh *et al.* 1991). The HP1 class is characterized mainly by glutamic acid rich block of six residues upstream of the chromobox. The third group, the Pc class, comprises the *Drosophila Pc* gene (Paro & Hogness 1991) itself and its putative murine homologue the *Mod3* gene (Pearce *et al.* 1992).

A search of the EMBL data base with the *CHD-1A* putative helicase domain (residues 451-911, Fig. 9) raises the identity between this and *CHD-1Y* to 55% in an overlap of 471 amino acids. There is also significant, but lesser identity to, the putative helicase motifs in the human (Okabe *et al.* 1992), and *S. cerevisiae* (Laurent *et al.* 1992) *SNF2* gene,

human (Muchardt & Yaniv 1993) and *Drosophila Brahma* (Tamkun *et al.* 1992), *S. cerevisiae* *NPS1/STH1* (Laurent *et al.* 1992, Tsuchiya *et al.* 1992), human excision repair protein *ECCR6* (Troelstra *et al.* 1992) and the *RAD54* (Emery *et al.* 1991) and *MOT1* (Davis *et al.* 1992) genes of *S. cerevisiae*. It should be noted that none of these latter genes contain a chromobox.

Only the four CHD genes show significant homology to the third motif, a DNA binding region identified by Delmas *et al.* (1993), whilst only *CHD-1A* and *CHD-1* have the three short basic HSDHR motif near the carboxy terminus, although this region is yet to be sequenced in *CHD-W*. The *CHD-1Y* gene apparently terminates before this point so does not share this motif. An extended discussion of the homology of the mouse *CHD-1* gene can be found in (Stokes & Perry 1995).

15 Isolation of *CHD-W*

Two, CC14 and CC4, of eight *CHD-1* related clones isolated from the 10 day chick embryo library using 349-4359nt of *CHD-1A* as a probe, overlap (Fig. 5) to provide the 1316bp of sequence given in Fig. 8. This is a sequence closely related to, but distinct from *CHD-1A*. Identity over the 1316bp overlap is 90.5% and 90.1% at the nucleotide and amino acid level respectively. An alignment of the putative translations of *CHD-1*, *CHD-1A* and *CHD-W* is given in Fig. 9. The amino acid identity between *CHD-1* and *CHD-1A* at 93.4% is marginally lower than that between that of *CHD-1* and *CHD-W*, 94.2%, over the same region

25 The 1335bp insert of CC4 was used at moderate stringency to probe a male/female, *PvuII* cut genomic blot featuring mouse, ostrich (*Struthio camelus*), chicken, bee-eater and hyacinth macaw (Fig. 13). Hybridization with the mouse and ostrich shows no evidence of any sex linkage, bands of the same size and equal intensity appearing in both
30 sexes. Hybridization with the ostrich is particularly strong, greater even

than with the cognate sequence in the chicken. This suggests that the genome size of the ostrich is considerably smaller than that of the chicken. It also demonstrates that the CHD genes cannot be used to sex the ostrich and, it is suggested, the other members of the ratites. There is no
5 evidence from further work (reported later) that this effect should occur in other Parvclasses of the birds (Sibley *et al.* 1988).

In all the bird species apart from the ostrich, hybridization occurs with two types of fragment some that are female unique and others that are shared between the sexes. In the chicken some of the latter are of
10 the same size as those hybridizing with the *CHD-1A* probe and result from cross hybridization under the conditions of low stringency that we employed. When probed with the CC4 sequence it is clear that hybridization with the female linked fragments is far stronger, at least in the chicken than with the shared fragments (bear in mind, also, that the female
15 chicken only has a single dosage of the W linked gene). This indicates that CC4 is W linked and represents part of *CHD-W*.

The DNA contained in the Southern blot of the male and female chickens probed in Fig.13 contained identical amounts of DNA. However, examination shows that the shared bands are twice as strong in
20 males (ZZ) as they are in females (WZ). The only way this could have happened is if the *CHD-1A* gene is Z linked. It is suggested this is the case in all birds.

Sex identification with PCR on dried and limited DNA in a Spix's 25 Macaw

The first test was devised to sex DNA extracted from the feathers of the last wild Spix's Macaw. This was the rarest bird on the planet and needed to be sexed so a mate could be selected from the 31 captive birds that remained. The test presented two problems. The first
30 was extracting DNA from feathers the second providing a test that would

work.

The procedure was published in Griffiths & Tiwari (1995) which covers the extraction of the DNA. The second test was to provide DNA from a Hyacinth Macaw which would yield data to allow construction of primers. A IFIX II library was provided by Stratagene and this was
5 probed with the insert of the *CHD-1A* clone Z6 (-227-5302 Fig. 6) at moderate stringency. This provided 7 positive clones (A1, A2, A7, A8, A13, 1.2 and 5C). The inserts were extracted cut with *MboI* and subcloned into the *BamI* cut pUC18. This sublibrary was probed again with the Z6
10 insert but this time at high stringency. The A12.3 subclone hybridized. This was sequenced and contained 111bp which is aligned to the chicken and mouse CHD genes in Fig 14. The similarity of this fragment to the chicken *CHD-W* suggested this was the Hyacinth Macaw homologue of the W chromosome located gene.

15 The data from A12.3 supplied information for the design of the primers required. It also provided evidence that the CHD sequences were sufficiently conserved in this region that a single set of primers could be designed to amplify both genes. Three primers, P1, P2 and P3, were designed to allow seminested PCR (Fig. 14). This technique allowed
20 amplification of a 104bp region of both *CHD-W* and *CHD-1A* from DNA that was available from two captive Spix's Macaws of known sex. In each sex the PCR products were of the same size but sequence determination revealed that the *CHD-W* derived PCR product possessed a *DdeI* restriction enzyme site which was lacking in the *CHD-1A* product. Thus
25 PCR amplification and *DdeI* cleavage of male Spix's Macaw DNA yields a only single product of 104 base pairs (bp), whilst from female DNA two products are apparent, one of 104bp and one of 73bp. The presence of the *CHD-1A* product in both sexes acts as a control to ensure the PCR amplification has been successful (Fig 15 & 16).

30 DNA was extracted from feathers moulted by the wild Spix's

Macaw using a technique devised for the purification of ancient DNA (Thomas & Pääbo 1993). The PCR-based test described above was used to demonstrate that *CHD-W* was not present in the sample (see Fig 16). This confirmed that the wild bird is male. A female Spix's macaw was released in March 1995 as a prospective mate.

Sex identification with PCR on a variety of birds

Birds can be sexed from DNA by showing the presence (female: ZW) or absence (male: ZZ) of the female specific W chromosome. At the molecular level this is carried out by the recognition of a W-linked marker. This can only be done after a W chromosome DNA marker is identified in the avian species. The test developed for the Spix's Macaw used *CHD-W* as a W linked marker. The data collected in designing this test suggested that this method may work to sex a variety of birds.

If the same test is to work on other bird species then two criteria must be met. The first is whether the PCR primers will amplify both *CHD* genes in other bird species. The Spix's Macaw test used the tiny amounts of DNA extracted from feathers so a seminested PCR was required. This used 3 primers which are aligned to the Mouse and Chicken *CHD* nucleotide sequences in Figure 14. The primer sites are highly conserved, there is no difference between the chicken genes and a solitary difference between the Mouse and Chicken in the 5' region of the P2 site. Theoretically, the primers should anneal to other bird species and, if a reasonable amount of DNA is available (>50ng), a single pair of primers should provide sufficient amplification.

A second requirement for the test is that the PCR products can be separated using a restriction endonuclease. In the Spix's Macaw the *DdeI* enzyme cuts *CHD-W* but not *CHD-1A*. Figure 14 shows that this discrimination would also occur in the Chicken. However, the *DdeI* cutting site CTNAG is not present in the *CHD-1A* of Spix's Macaw (CTNGG) nor

the Chicken (CANAG) for different reasons. This suggests that the *DdeI* site is open to mutation so this form of discrimination is unlikely to be conserved. Other discriminatory sites are available: *DdeI* and *MaeII* sites are unique to *CHD-W* and the *HaeIII*, *MboII* and *XhoI* sites to *CHD-1A* and
5 can be considered the first option. If these fail the *CHD-W* and *CHD-1A* PCR fragments can be cloned and sequenced so discriminatory sites can be discovered.

The theory we have presented suggests that a sexing test based on both avian CHD genes should work on many other bird species.
10 Does this work in practice? The birds selected for trial are from across the avian class: Chicken (5 individuals), Marbled Murrelet (18), Kestrel (8), Marsh Harrier (28), Bee-eater (4), 1 pair of six species of Strigidae Owls from different genera (see Methods), Starling (5) and African Marsh Warbler (5).

15 The primers amplify a PCR product of the predicted size in all of the birds using primers P2 and P3 on 50-100ng of genomic DNA extracted from blood. Figure 17 illustrates this for 3 bird species but also includes amplification from human DNA. This shows that tests using P2 and P3 are open to human DNA contamination so appropriate precautions
20 must be taken.

The *HaeIII* restriction enzyme cut the *CHD-1A* fragment alone in all 13 species (Fig 17) and, from the sequence data, would also have worked on the Spix's Macaw (Fig 16). Figure 17 shows that the *CHD-1A* in males is cut into two fragments (45bp, 59bp) which are not easily visible on
25 the gel. In females *CHD-W* is uncut by *HaeIII* so remains at 104bp. The discrimination using *HaeIII* provided correct sex identification in all individuals.

Discussion

The CHD genes

The female specific great tit probe GT-W was described by Griffiths and Tiwari (1993) as a means of identifying sex in this species.

5 The results presented here suggest this sequence represents part of a intron in a W linked gene. By moving downstream from this sequence it has been possible to isolate a putative exon from a gene that we have named *CHD-W* due to its close sequence identity to the mouse *CHD-1* gene (Delmas *et al.* 1993) and its W location.

10 Using the *CHD-W* fragment we attempted to isolate a similar, W linked sequence that Southern blot analysis had shown was present in the chicken. From several clones a 6606bp cDNA sequence was assembled but although it has close sequence identity to the great tit *CHD-W* fragment Southern blot analysis shows it is not located on the W
15 chromosome. This second gene was termed *CHD-1A* (A = avian). This blot shows a second gene closely related to *CHD-1A* is W located. This sequence could not be cloned from a stage 10-12 chick cDNA library although 19 *CHD-1A* clones were isolated. However, two clones yielding 1347bp of a second CHD gene were isolated along with a further 14
20 *CHD-1A* clones from a day 10 chick cDNA library. Southern blot analysis showed that this second clone was W chromosome derived and so represents *CHD-W*. Attempts are underway to isolate the remainder of *CHD-W*.

Southern blots of a variety of bird species showed that
25 *CHD-W* is W chromosome linked in all birds except the ostrich. This suggests that the gene is sex linked throughout the class with the exception of the primitive ratites, which the ostrich represents, where it appears to be autosomally located.

An alternative explanation is that the *CHD-W* is in fact W
30 linked in ratites but occurs in a region of the W chromosome which still

recombines with the Z chromosome. If *CHD-1A* were Z linked, then recombination between Z and W linked copies of CHD would maintain their sequence identity resulting in the apparently autosomal location indicated by the Southern blot. A mammalian example would be the *MIC2* and *STS* genes that are located in the pseudoautosomal region of the Y chromosome (Ellis & Goodfellow 1989) and would give analogous results to those observed here.

Two lines of evidence support this alternative hypothesis. The first is that the Southern blot analysis suggests that *CHD-1A* is Z linked in non-ratites which would make the chromosomal location of the CHD-genes consistent throughout the class. Hybridization of *CHD-1A* to genomic blots is apparently stronger to fragments from male birds which would result from this sex having two copies of any Z linked gene in comparison to a single copy in the female (this result is not clear cut and requires confirmation by chromosomal *in situ*). The second line of evidence is that the sex chromosomes of the ratites are not morphologically differentiated as is the case with other birds (Christidis 1990). Morphological similarity suggests recombination still occurs between extensive regions of the ratite Z and W which may include the CHD genes and so produce the pattern of hybridization observed.

Although we have yet to clone the whole of *CHD-1A* the 6606bp sequenced so far shows a close identity to the mouse *CHD-1* gene over the putative coding region. It also includes all four features identified by Delmas *et al.* (1993) as having possible functional significance. This includes a chromodomain, a helicase, a DNA binding motif and a basic, five amino acid motif that is repeated three times (Fig. 9). The similarity of the sequence derived thus far from *CHD-W* to that of *CHD-1* and *CHD-1A* suggest it will be of similar length and possess these same motifs. We have also identified an alternatively spliced form of *CHD-1A* and *CHD-W* which has a similar adenine rich motif inserted at an identical point

(4327/4328nt *CHD-1A* and 1316nt *CHD-W*). The exact form of these alternative mRNAs is yet been elucidated. It is interesting to note that we obtained no clones that spanned these breakpoints which contained this additional motif; the sequence was built up from partial sequences derived from either 5' or 3' termini of different clones. Delmas *et al.*, (1993) produced a mRNA Northern blot probed with fragments of *CHD-1* occurring 5' to this breakpoint and discovered an mRNA species of about 4kb. This would correspond to a species cleaved near this insertion point. What purpose this would serve is unknown. Moreover the putative yeast homologue of *CHD*, *CHD-1Y*, which was identified from amino acid identity to *CHD-1A* from the genomic sequence on the EMBL database does not apparently have a similar motif. This is suggested because the *CHD-1Y* sequence was derived from a genomic clone which would allow the identification of any such sequence were it to be spliced in the normal manner.

The significance of the four functional domains found in the *CHD* genes will be discussed in turn. The first, the carboxy-terminal trimer repeat of five basic amino acid residues, has no known function and is not shared by any other sequences from the EMBL database. Furthermore, the *CHD-1Y* gene which is truncated by a little over 200 amino acid residues in comparison to *CHD-1* and *CHD-1A* does not contain this motif.

The second functional domain was identified by Delmas *et al.* (1993) as having sequence selective DNA binding capacity. Whether this is highly specific or just to A+T rich regions was not established. They also noted that this domain contains Lys-Arg-Pro-Lys-Lys and Arg-Gly-Arg-Pro-Arg motifs which enable genes like *HMG-1*, *D1* and *Engrailed* to bind in the minor groove of A+T rich DNA.

A third functional motif is located towards the N-terminus of the *CHD*-protein and is termed the chromodomain [Chromatin Organization Modifier; Paro, 1990 #459]. This is a highly conserved domain of between

37-50 amino acids that has been shown to be represented in the genomes of plants, nematodes, insects and vertebrates (Singh *et al.* 1991). Several chromobox genes have been isolated from human, mouse and *Drosophila* and have been divided into the *polycomb* (Pc) class and the
5 heterochromatin protein-1 (HP1) class on the basis of related structure (Pearce *et al.* 1992)). The CHD-genes have a distinct form of the chromobox characterized by close homology between yeast and vertebrate forms in the 5' half of the box itself but extending a further 17 residues downstream. These differences indicate that this form of the chromobox
10 defines a third subgroup the CHD class

The *Pc* gene forms one of a eponymously named group (Pc-g) of about 12 genes defined through homeotic mutants in *Drosophila* that prevent fixation and maintenance of a determined state. They act as transcriptional repressors of homeotic genes, notably of the antennapedia
15 complex (ANT-C; Paro, 1990). Members of the ANT-C and the other major group of *Drosophila* homeotic genes, the bithorax complex (BX-C), are responsible for defining segmental identity during development (Kaufman *et al.* 1980, Lewis 1978). Initially, their expression patterns are designated by early acting maternal and segmentation genes (see 4,6,7 Kennison).
20 However, these maternal genes are only transiently expressed. During the later stages of development their role as transcriptional activators is adopted by an assemblage of genes including the trithorax group (Trx-g), whilst many of their repressive effects are assumed by the Pc-g (Kennison 1993).

25 The *polycomb* (Pc) gene itself is perhaps the best studied member of the Pc-g. Zink and Paro (1989) used *Pc-B*-galactose fusion proteins to show that it binds to around 100 different sites on the polytene chromosome including loci where other members of the Pc-g are located. Any disruption of the chromodomain abolishes the specificity of this
30 reaction (Messmer *et al.* 1992). However, the Pc-g protein appears to lack

any type of endogenous DNA binding capacity so it is thought that it acts as part of a protein complex with other components that are responsible for the site specific DNA binding (Paro 1990).

The repressive effects of the Pc-g are thought to be the result of chromatin compaction. In other words, the DNA is packaged into heterochromatin to prevent or reduce the expression of functional genes (Paro 1990). This is a mechanism related to position effect variegation (PEV; (Henikoff 1990)), to dosage compensation in mammals which sees the complete heterochromatization of one of the female's X chromosomes and possibly to gene imprinting whereby the expression of maternally and paternally inherited alleles differs (Peterson & Sapienza 1993). The links with PEV have recently been substantiated in that HP1, a recognized modifier of PEV, and *Pc* both contain chromodomains (Paro & Hogness 1991). Like the *Pc* protein, HP1 appears to form part of a structural complex that transforms euchromatin to heterochromatin. Furthermore, both PEV and the repressive effects of *Pc* are passed, in a clonal manner, to daughter cells ((Henikoff 1990, Struhl 1981); a characteristic also of gene imprinting.

With the CHD-type gene containing both a DNA binding motif and a chromobox it may appear reasonable to suggest that they encode repressors with an endogenous, site selective DNA binding system. However, CHD genes contain a further functional motif that is structurally related to the *Helicases*. The sequence identity is closest to the yeast *SNF2/SWI2* (Abrams *et al.* 1986) and *Drosophila Brahma* genes (Tamkun *et al.* 1992), both of which are transcriptional *activators*. Indeed, *Brahma* is part of the Trx-g which are considered direct antagonists to the Pc-g. Other genes which contain more distantly related *Helicase* domains are involved in DNA repair and chromatid separation during mitosis (Laurent *et al.* 1993, Sung *et al.* 1993).

The *SWI2* gene product has been shown to enhance the

transcription of other genes probably as part of a complex that includes *SWI1*, *SWI3*, *SNF5*, *SNF6* and in conjunction with gene specific DNA binding proteins (Laurent *et al.* 1991, Peterson & Heskowitz 1992). A mode of action strikingly similar to that of *Pc*.

5 Although it remains to be formally demonstrated that *SWI2* is a helicase, it does have close structural similarities with proven *Helicase* genes and also possesses the required DNA stimulated ATPase activity (Laurent *et al.* 1993). Laurent *et al.*, go on to postulate that the *SWI2* containing complex may act by two mechanisms acting either separately or
10 in conjunction. In the first they envisage helicase mediated DNA melting to allow the egress of RNA polymerase II. Alternatively *SWI2* could allow chromatin remodelling, in effect overcoming any inhibitory packaging of the DNA and so enhancing transcription.

 The juxtaposition of a *Helicase* and a chromodomain within
15 the same gene presents a paradox that may challenge the perceived roles of the two motifs. A simple explanation is that alternative splicing could remove one or other of these domains prior to translation. However, there is little support for this idea from the work of ourselves or Delmas *et al.*, (1993).

20 An alternative explanation could be due to our lack of real knowledge about the function of the chromobox. Whilst it is well² established that *Helicases* do disassociate DNA and so facilitate transcription (Matson & Kaiser-Rogers 1990), the role of the chromodomain in repression is based on more circumstantial evidence.
25 *Pc*, as we have seen, does not bind DNA itself although mutations in the chromobox prevent the formation of site specific complexes. It is possible that the chromodomain is involved more in maintaining the structural integrity of the repressive complex than in the repressive mechanism itself. Based on this supposition, the CHD-protein may form a different type of
30 complex able to bind at a site dictated or influenced by its own binding

domain and activate these loci via helicase activity.

While both this scenario is speculative it is probable that CHD-type genes are active during development and are able to bring about heritable changes in transcription. The presence of an endogenous DNA binding domain suggests it has fewer targets than *Pc*, for example, which could form part of several different active complexes. With *CHD-W* being confined to the W chromosome is likely to have a role in some aspect of female development and we suggest this may be critical to the determination of gender. In support this hypothesis we were unable to find any *CHD-W* clones in a library constructed prior to sex determination which occurs at day 7 (Lutz-Ostertag 1954) but were able to isolate two clones from a smaller pool of candidates at day 10. This suggests that the expression of *CHD-Y* may occur at a time consistent with its having a sex determining role.

If *CHD-W* alone or in conjunction with *CHD-1A* causes sex determination in birds then several potential mechanisms are plausible.

- (1) In the simplest scenario *CHD-Y* may act as a simple trigger like *SRY* (Koopman 1993) to either cause expression or repression of downstream genes in order initiate testis development.
- (2) *CHD-W* may interact with other autosomal or Z linked genes whereby the dosage of *CHD-W* in comparison these other factors causes initiates development down the male or female pathways.

A more complicated scenario is if *CHD-W* acts in together with *CHD-1A* to cause sexual differentiation. Different mechanisms could operate depending whether *CHD-1A* turns out to be Z linked as we suspect or autosomal.

- (3) If *CHD-1A* is Z linked, then male birds get two doses of the *CHD-1A* expression product to one in female birds. Perhaps the 1:1 ratio of functionally distinct *CHD-1A* and *CHD-W* products is what initiates female development whilst a double dosage of *CHD-1A* results in males.

(4) Alternatively, just the single dosage of Z linked *CHD-1A* product could result in female development and expression of *CHD-W* only occurs after sexual differentiation to equalize dosages of functionally similar proteins.

5 (5) If *CHD-1A* is autosomal however, it could be envisaged that *CHD-1A* and *CHD-W* are functional homologues and the three doses in females (AAW) is required to promote female development, whilst the double dosage in males (AA) causes the differentiation of the testis and the development of the male phenotype.

10 The evidence from aneuploid chickens discussed in the introduction, does suggest that the mechanism that does operate involves some degree of dosage dependence which tends to exclude mechanism (1). However the similarity of *CHD-W* to *HP1*, the *Pc* protein and other transcriptional modifiers that act through chromatin remodelling show that
15 the expression of this type is crucially dependent on dosage (Locke *et al.* 1988). With the different dosages of gene product and/or potential target sites that aneuploids possess it may be that analysis of these type of mutants has, thus far, served to confuse the issue.

20 Sex Identification

The first W-chromosome linked DNA was isolated by Tone *et al.* (1982) from the Chicken. Since then, a number of other W-linked avian sequences have been discovered (e.g. Griffiths, 1990; Rabenold, 1991; Griffiths, 1993). In all but one case, described later, these DNA fragments
25 appear to be non-functional repeats. For instance, the related *XhoI* and *EcoRI* fragments in Chicken may comprise 70-90% of the W chromosome (Saitoh *et al.* 1991). This repeat and others in the Lesser Black-backed Gull (*Larus fuscus*) can be used to sex birds by the rapid dot blotting technique (Griffiths & Holland 1990). Other less repetitive W chromosome
30 markers can be used to sex birds either by probing Southern blots

(Rabenold *et al.* 1991) or through the use of PCR (Griffiths & Tiwari 1993).

The major problem with all non-functional W-linked DNA is the speed with which they evolve. The chicken *XhoI* repeat is fairly typical. Through low stringency hybridization to a Southern blot it can be used to
5 sex the Turkey (*Meleagris gallopavo*) and the Pheasant (*Phasianus versicolor*, Saitoh *et al.* 1991). These bird species are closely related to the Chicken by being members of the family Phasianidae. By contrast, the functional *CHD-W* region described here is 96% (3/67 Fig 3) identical
10 between Chicken and Spix's Macaw and this only drops to 86% between the Chicken *CHD-W* and the Mouse *CHD1* (15/110 Fig 3). This level of conservation means that the chicken *CHD-W* probe can be used on Southern blots to sex birds from all over the class Aves.

The only exception to the non-functional avian W-linked sequences is *DZWM1* which is a putative gene, cloned from a cDNA turkey
15 library. Like *CHD-W* this gene appears to be sex linked in many bird species. Unfortunately, so little information has been published in the papers that describe *DZWM1* that the nature of the gene remains unknown (Dvorák *et al.* 1992, Halverson 1990, Halverson & Dvorák 1993).

For sexing large numbers of birds Southern blot analysis is
20 slow and expensive. The technique that we have used is based on a PCR using P2 and P3 primers followed by a *HaeIII* digestion of the of the amplified product. The digestion distinguishes between the *CHD-W* product which is uncut and the *CHD-1A* which is cut. The technique will work to sex a range of bird species that span the class Aves. The primers
25 target a highly conserved region so are likely to be 'universal' to the birds but the discriminatory *HaeIII* site which cuts *CHD-1A* but not *CHD-W* shows no real reason to be conserved. If *HaeIII* does fail to be discriminatory other cutting sites have been suggested or the *CHD-W* and *CHD-1A* PCR products can easily be sequenced to look for an alternative.
30 Alternatively, the different nucleotide sequence of the amplified *CHD-W*

and *CHD-1A* suggests that the two PCR products would be separable on an agarose gel of around 3% or a non-denaturing acrylamide gel. This would remove the need for a cutting enzyme and may well make the sexing technique more easy to use.

5 The CHD based test appears to be fairly solid but the chances of a peculiar mutation in some bird species is not impossible. Cases concerning *SRY/Sox3* genes on the sex chromosomes in mammals supports this claim. In two species of the vole *Ellobius* males have neither a Y chromosome nor an *SRY* gene (Just *et al.* 1995). In a second case,
10 four species of *Akodon*, the Mole Vole, have 15-40% of fertile females with XY chromosomes and an *SRY* gene (Bianchi *et al.* 1993). These examples are particularly peculiar in that the *SRY* gene is accepted as the gene that determines sex throughout the mammals. In neither case would the detection of *SRY* reliably inform you of the animals sex.

15 These examples from the Muridae may never occur with the CHD genes of birds. However, it does suggest that sex identification by the amplification of *CHD-W* and *CHD-NW* should always be validated by a test on several individuals in a new species before it is applied. Despite this warning, the use of the test described here or by other means using
20 the *CHD-W* or *CHD1A*, these genes provide a method to sex most bird species.

Table 1. Sex of domestic fowl with normal and abnormal chromosome complements (from McCarrey & Abbott (1979) and Crew (1954)).

Chromosome complement	Phenotype
AAZZ	Male
AAZW	Female
AAZZW	Male?
AAZZZ	Male
AAAZZZ	Male
AAAZZW	Intersex/male

Figure Legends

Figure 1. The DNA sequence of the pGT-W insert.

Figure 2. A map of the 9.6kb insert of the IFixII clone isolated from the great tit using pGT-W. pGT1.7 and pGT8 are the two *EcoRI* subclones into which the fragment was divided. The broken line corresponds to the region with absolute sequence identity to the pGT-W insert. The position of the region with identity to the mouse *CHD-1* gene is indicated.

Figure 3. An alignment of 123bp fragment of the great tit (GT) *CHD-W* gene in pGT8 with the autosomal/Z located chicken (C) *CHD-1A* the chicken *CHD-W* gene and bases 3855-3977 of the mouse (M) *CHD-1* gene. An alignment of the deduced amino acid sequence is also given.

Figure 4. The section of pGT8 that hybridized to a female specific fragment of 3.1kb in the chicken. This probe was also used to screen the chicken cDNA library. The hatched line represents the female specific great tit motif shown in Fig. 3.

Figure 5. The complete nucleotide sequence of CHD-1A as defined by the clones Z4, Z6 and Z11. Two asterisks underlie the position where part of the sequence illustrated in Fig7 is spliced onto the 5' or 3' ends of a proportion of the clones isolated. The ATG at nucleotide 228 is the start codon whilst TAA at 5388 is the stop codon.

Figure 6. The strategies used to determine the nucleotide sequence of *CHD-1A* and *CHD-W* given in Fig. 5 and Fig. 8. The top line represents the mouse clone given by (Delmas et al. 1993). The three 'Z' clones of *CHD-1A* and the 'CC4' and 'CC14' clones of *CHD-W* were derived from either a stage 10-12 or a 10 day chick cDNA library respectively. Arrows indicate the direction of sequence determination. Note Z6 actually ran from -227 to 69. These nucleotides were determined and are found in Fig 5

Figure 7. A composite nucleotide sequence and putative translation of the motif that is found spliced to a proportion of the 5' or 3' termini of *CHD-1* clones or the 3' end of the *CHD-W* clone CC14. The portion attached to the CC14 sequence is incomplete.

Figure 8. A partial nucleotide sequence of *CHD-W* as defined by the clones CC4 and CC14.

Figure 9. An alignment of the deduced amino acid sequences of the chicken (C) *CHD-1A* and *CHD-W* with the mouse (M) *CHD-1*. With gaps introduced to maximize alignment they show a sequence identity of 91.6% over 1365 residues. The \$ sign indicates start and stop codons. Boxed sections are the chromodomain (C), Helicase (H), and the region containing the DNA binding domain (B) identified by Delmas et al., (1993). A trimer repeat of a basic HSDHR motif is underlined. A* denotes residue identity and . similarity.

Figure 10. An alignment of the deduced amino acid sequences of *CHD-1A* and *CHD-1Y* a putative yeast homologue of the chicken gene identified through a search of the EMBL data base. With

gaps introduced to maximize alignment they show a sequence identity of 37.7% over 1538 residues. | indicates identity and : conservative substitution.

Figure 11. Comparison of 9 chromodomain sequences.

- 5 Vertical lines indicate the extent of the chromodomain as defined by Paro & Hogness (1991). The top three sequences represent the CHD class of chromodomain to add to the HP1 class and Pc class][;-l08k9ouygytrdevz as defined by Pearce et al. (1992). The first letter of each annotation indicates the animal of origin: C, chicken; M mouse; D, Drosophila; H, human; Y, *S. cerevisiae* whilst the remainder identifies the gene type. The yeast gene is a possible CHD homologue identified by its close identity to the vertebrate forms. * indicates sequence identity within the groups and ^ identity between all nine sequences. _ indicate amino acid residues inside and downstream of the motif that are characteristic of the CHD class chromobox.
- 10
- 15

- Figure 12. Genomic Southern blots of DNA from male and female chickens and lesser black-backed gulls digested with PvuII and probed with a 433bp HindIII/Sac fragment of pGT8 (Fig 4.) at moderate stringency. Hybridization with female linked fragments and fragments common to both sexes can be observed in both species. Numbers give approximate sizes in kilobases.
- 20

- Figure 13. Genomic Southern blots of DNA from male (M) and female (F) mice, ostrich, chicken, bee-eater and hyacinth macaw probed with the 1335bp insert of CC4 at moderate stringency.
- 25 Hybridization with mouse and ostrich is with fragments shared by both sexes whilst the non-ratite birds show additional hybridization to female specific fragments. In these latter species, the signal from female linked hybrids is stronger than with autosomal/Z linked fragments indicating that the probe is derived from the W chromosome. Numbers give approximate sizes in kilobases.
- 30

- 43 -

Figure 14. The nucleotide sequence of part of a single CHD1 gene isolated from the Mouse and the homologous genes from the Chicken, Hyacinth (A12.3 subclone) and Spix's Macaw all arranged as putative codons. Dashes denote nucleotides shared with the Mouse sequence. The primers designed are shown on the diagram. An arrow head indicates a non-synonymous mutation in the Spix CHD-W. The *DdeI* (CTNAG) and *HaeIII* (GGCC) sites are underlined.

Figure 15. The technique of PCR sex identification in the Spix's Macaw. Semi-nested PCR amplification is carried out on both sexes with the primers P2/P3 then P1/P2 to provide products of identical sizes in both sexes. The products are then cut with restriction enzyme *DdeI* which cuts only the CHD-W product from the female. The cut products are run on a visigel and the difference between the sexes can be visually detected. See Fig 17 for an example.

Figure 16. *DdeI* restricted PCR products demonstrating that remaining wild Spix's Macaw is male. Lane 1. the wild bird 2. negative extraction control 3. known male 4. known female. The larger fragment is of 104 bp and the female W-chromosome specific fragment of 73 bp.

Figure 17. Sex identification in the Marsh Harrier (MH), Chicken (C) and African Marsh Warbler (AMW) carried out using an identical reaction. For each species genomic DNA of male and female birds was subject to PCR with primers P2 and P3 and the product of 110bp is visible in lanes 1 and 2. In lane 3 the entire male PCR product, amplified from *CHD-1A*, has cut into two parts with *HaeIII* (65bp, 45bp). In females, lane 4 this *HaeIII* cut product is also present but the *CHD-W* product remains uncut so the sex can be identified. The 'Kb' lane contains a '1Kb DNA ladder' (BRL), the 'H' lane is PCR reaction with P2 and P3 carried out on human genomic DNA and -ve lane contains a negative PCR reaction.

References

- Abrams, E., Neigeborn, L. & Carlson, M. 1986. Molecular analysis of SNF2 and SNF5, genes required for expression of glucose-repressible genes in *Saccharomyces cerevisiae*. Mol. Cell. Biol. 6: 3643-3651.
- Baverstock, P. R., Adams, M., Polkinghorne, R. W. & Gelder, M. 1982. A sex-linked enzyme in birds - Z-chromosome conservation but no dosage compensation. Nature 296: 763-766.
- 10 Bianchi, N. O., Bianchi, M. S., Bailliet, G. & de la Chapelle, A. 1993. Characterization and sequencing of the sex determining region Y gene (SRY) in Akodon (Cricetidae) species with sex reversed females. Chromosoma 102: 389-395.
- Charlesworth, B. 1991. The evolution of sex chromosomes. 15 Science 251: 1030-1033.
- Christidis, L. 1990. Animal cytogenetics 4: Chordata 3; B, Aves. Berlin, Gebrüder Borntraeger.
- Crew, F. A. E. 1954. Sex determination. New York, John Wiley and Sons, Inc.
- 20 Davis, J. L., Kunisawa, R. & Thorner, J. 1992. A presumptive *Helicase* (MOT1 gene product) affects gene expression and is required for viability in the yeast *Saccharomyces cerevisiae*. Mol. Cell. Biol. 12: 1879-1892.
- Delmas, V., Stokes, D. G. & Perry, R. P. 1993. A mammalian 25 DNA binding protein that contains a chromodomain and an SNF2/SW12-like *Helicase* domain. Proc. Natl. Acad. Sci. USA 90: 2414-2418.
- Dretzen, G., Bellard, M., Sassone-Corsi, P. & Chambon, P. 1981. A reliable method for the recovery of DNA fragments from agarose and acrylamide gels. Anal. Biochem. 112: 295-296.
- 30 Dvorák, J., Halverson, J. L., Gulick, P., Rauen, K. A., Abbott,

U. K., Kelly, B. J. & Shultz, F. T. 1992. cDNA cloning of a Z- and W-linked gene in gallinaceous birds. *J. Heredity* 83: 22-25.

Ellis, N. & Goodfellow, P. N. 1989. The mammalian pseudoautosomal region. *Trends Genet.* 5: 406-410.

5 Emery, H. S., Schild, D., Kellogg, D. E. & Mortimer, R. K. 1991. Sequence of RAD54 a *Saccharomyces cerevisiae* gene involved in recombination and repair. *Gene* 104: 103-106.

Foster, J. W., Brennan, F. E., Hampikian, G. K., Goodfellow, P. N., Sinclair, A., Lovell-Badge, R., Selwood, L., Renfree, M. B., Cooper,
10 D. W. & Graves, J. A. M. 1992. Evolution of sex determination and the Y chromosome: *SRY*-related sequences in marsupials. *Nature* 359: 531-533.

Gietz, R. D. & Prakash, S. 1988. Cloning and nucleotide sequence analysis of the *Saccharomyces cerevisiae* RAD4 gene required for excision repair of UV-damaged DNA. *Gene* 74: 535-541.

15 Grant, S. G. & Chapman, V. M. 1988. Mechanisms of X-chromosome regulation. *Ann. Rev. Genet.* 22: 199-233.

Graves, J. A. M. 1987. The evolution of mammalian sex chromosomes and dosage compensation: clues from marsupials and monotremes. *Trends Genetics* 3: 252-256.

20 Griffiths, R. 1991. The isolation of conserved DNA sequences related to the human sex-determining region Y gene from the lesser black-backed gull (*Larus fuscus*). *Proc. R. Soc. Lond. B.* 244: 123-128.

Griffiths, R. & Holland, P. W. H. 1990. A novel avian W chromosome DNA repeat sequence in the lesser black-backed gull (*Larus*
25 *fuscus*). *Chromosoma* 99: 243-250.

Griffiths, R. & Kelsey, M. 1995. A mate for the last macaw. *Daily Telegraph* 43,544: 24.

Griffiths, R. & Tiwari, B. 1993. The isolation of molecular genetic markers for the identification of sex. *Proc. Natl. Acad. Sci. USA* 90:
30 8324-8326.

Griffiths, R. & Tiwari, B. 1995. Sex of the last wild Spix's macaw. *Nature* 375: 454.

Halverson, J. L. 1990. Avian sex identification by recombinant DNA technology. *Proc. Assoc. Avian Vet.* 256-262.

5 Halverson, J. L. & Dvorak, J. 1993. Genetic control of sex determination in birds and the potential for its manipulation. *Poultry Sci.* 72: 890-896.

Henikoff, S. 1990. Position-effect variegation after 60 years. *Trends Genet.* 6: 422-426.

10 Hodgkin, J. 1992. Genetic sex determination mechanisms and evolution. *BioEssays* 14: 253-261.

James, T. C. & Elgin, S. C. R. 1986. Identification of a nonhistone chromosomal protein associated with heterochromatin in *Drosophila melanogaster* and its gene. *Mol. Cell. Biol.* 6: 3862-3872.

15 Jost, A. 1947. Sur les effets de castration precoce de le embryo male du lapin. *C. R. Soc. Biol.* 141: 126-129.

Just, W., Rau, W., Vogel, W., Akhverdian, M., Fredga, K., Graves, J. A. M. & Lyapunova, E. 1995. Absence of *SRY* in species of the *Vole Ellobius*. *Nature Genetics* 11: 117-118.

20 Kaufman, T. C., Lewis, R. & Wakimoto, B. 1980. Cytogenetic analysis of chromosome 3 in *Drosophila melanogaster*: the homeotic gene complex in polytene chromosome interval 84A-B. *Genetics* 94: 115-133.

Kay, G. F., Ashworth, A., Penny, G. D., Dunlop, M., Swift, S., Brockdorff, N. & Rastan, S. 1991. A candidate spermatogenesis gene on the mouse Y chromosome is homologous to ubiquitin-activating enzyme. 25 *Nature* 354: 486-489.

Kennison, J. A. 1993. Transcriptional activation of *Drosophila* homeotic genes from distant regulatory elements. *Trends Genet.* 9: 75-79.

Koopman, P. 1993. Analysis of gene expression by reverse transcriptase - polymerase chain reaction. In: C. D. Stern & P. W. H. 30

Holland (eds). Essential developmental biology: 233-242. Oxford, IRL Press.

Koopman, P., Gubbay, J., Vivian, N., Goodfellow, P. & Lovell-Badge, R. 1991. Male development of chromosomally female mice
5 transgenic for *SRY*. *Nature* 351: 117-121.

Laurent, B. C., Treich, I. & Carlson, M. 1993. The yeast *SNF2/SWI2* protein has DNA stimulated ATPase activity required for transcriptional activation. *Denes & Dev.* 7: 583-591.

Laurent, B. C., Treitel, M. A. & Carlson, M. 1991. Functional
10 interdependence of the yeast SNF2, SNF5, and SNF6 proteins in transcriptional activation. *Proc. Natl. Acad. Sci. USA* 88: 2687-2691.

Laurent, B. C., Yang, X. & Carlson, M. 1992. An essential *Saccharomyces cerevisiae* gene homologous to SNF2 encodes a *Helicase* related gene. *Mol. Cell. Biol.* 12: 1893-1902.

15 Lewis, E. B. 1978. A gene complex controlling segmentation in *Drosophila*. *Nature* 276: 565-570.

Locke, J., Kotarski, M. A. & Tartof, K. D. 1988. Dosage-dependent modifiers of position effect variegation in *Drosophila* and a mass action model to explain their effects. *Genetics* 120: 181-190.

20 Lutz-Ostertag, Y. 1954. Contribution a l'etude canaux de muller chez l'embryon d'oiseaux. *Bull. Biol. Fr. Belg.* 88: 333-412.,

Matson, S. W. & Kaiser-Rogers, K. A. 1990. DNA helicases. *A. Rev. Biochem.* 59: 289-329., J. R. & Abbott, U. K. 1979. Mechanisms of genetic sex determination, gonadal sex differentiation, and germ-cell
25 development in animals. *Adv. Genet.* 20: 217-290.

Messmer, S., Franke, A. & Paro, R. 1992. Analysis of the functional role of the *Polycomb* chromo domain in *Drosophila melanogaster*. *Genes & Dev.* 6: 1241-1254.

Mitchell, M. J., Woods, D. R., Tucker, P. K., Opp, J. S. &
30 Bishop, C. E. 1991. Homology of a candidate spermatogenic gene from the

mouse Y chromosome to the ubiquitin-activating enzyme E1. *Nature* 354: 483-486.

Muchardt, C. & Yaniv, M. 1993. A human homologue of *Saccharomyces cerevisiae* *SNF2/SWI2* and *Drosophila* *brm* genes potentiates transcriptional activation by the glucocorticoid receptor. *EMBO J.* 12: 4279-4290.

Okabe, I., Bailey, L. C., Attree, O. F., Perkel, J. M., Nelson, D. L. & Nussbaum, R. L. 1992. Human and bovine homologues of *Saccharomyces cerevisiae* *SNF2/SWI2* a global activator of transcription in yeast. *Nucl. Acids Res.* 20: 4649-4655.

Page, D. C., Mosher, R., Simpson, E. M., Fisher, E. M. C., Mardon, G., Pollack, J., McGillivray, B., de la Chapelle, A. & Brown, L. G. 1987. The sex-determining region of the human Y chromosome encodes a finger protein. *Cell* 51: 1091-1104.

Paro, R. 1990. Imprinting a determined state into the chromatin of *Drosophila*. *Trends Genet.* 6: 416-421.

Paro, R. & Hogness, D. S. 1991. The *polycomb* protein shares a homologous domain with a heterochromatin-associated protein of *Drosophila*. *Proc. Natl. Acad. Sci. USA* 88: 263-267.

Pearce, J. J. H., Singh, P. B. & Gaunt, S. J. 1992. The mouse has a *polycomb*-like chromobox gene. *Development* 114: 921-929.

Peterson, C. L. & Heskowitz, I. 1992. Characterization of the yeast *SWI1*, *SWI2*, and *SWI3* genes, which encode a global activator of transcription. *Cell* 68: 573-583.

Peterson, K. & Sapienza, C. 1993. Imprinting the genome: Imprinted genes, imprinting genes, and a hypothesis for their interaction. *Annu. Rev. Genet.* 1993: 7-31.

Rabenold, P. P., Piper, W. H., Decker, M. D. & Minchella, D. J. 1991. Polymorphic minisatellite amplified on the avian W chromosome. *Genome* 34: 489-492.

Sachs, A. B., Bond, M. W. & Kornberg, R. D. 1986. A single gene from yeast for both nuclear and cytoplasmic polyadenylate-binding proteins: domain structure and expression. *Cell* 45: 827-835.

Saitoh, Y., Saitoh, H., Ohtomo, K. & Mizuno, S. 1991.

- 5 Occupancy of the majority of DNA in the chicken W chromosome by bent repetitive DNA sequences. *Chromosoma* 101: 32-40.

- Saunders, W. S., Chue, C., Goebel, M., Craig, C., Clark, R. F., Powers, J. A., Eissenburg, J. C., Elgin, S. C. R., Rothfield, N. F. & Earnshaw, W. C. 1993. Molecular cloning of a human homologue of
10 *Drosophila* heterochromatin protein HP1 using anti-centromere autoantibodies with anti-chromo specificity. *J. Cell Sci.* 104: 573-582.

Sibley, C. G., Ahlquist, J. E. & Monroe, B. L. 1988. A classification of the living birds of the world based on DNA-DNA hybridization studies. *Auk* 105: 409-423.

- 15 Singh, P. B., Miller, J. R., Pearce, J., Kothary, R., Burton, R. D., Paro, R., James, T. C. & Gaunt, S. J. 1991. A sequence motif found in *Drosophila* heterochromatin protein is conserved in animals and plants. *Nucl. Acids. Res.* 19: 789-794.

- Stokes, D. G. & Perry, R. P. 1995. The DNA-binding and
20 chromatin-localization properties of CHD1. *Mol. Cell Biol.* 15: 2745-2753.

Stoll, R., Rashedi, M. & Maraud, R. 1978. Sur l'hormone de regression MYllerienne, agent inducteur do testicule chez l'embryon de poulet. *Bull. Assist. Anant.* 176: 131-143.

- Struhl, G. 1981. A gene product required for correct initiation
25 of segmetal determination in *Drosophila*. *Nature* 293: 36-41.

Sung, P., Bailly, V., Weber, C., Thompson, L. H., Prakash, L. & Prakash, S. 1993. Human Xeroderma pigmentosum group D gene encodes a DNA *Helicase*. *Nature* 365: 852-855.

- Tagaki, N., Itoh, M. & Sasaki, M. 1972. Chromosome studies
30 in four species of ratitae (Aves). *Chromosoma* 36: 281-291.

Tamkun, J. W., Deuring, R., Scott, M. P., Kissinger, M.,
Pattatucci, A. M., Kaufman, T. C. & Kennison, J. A. 1992. Brahma: a
regulator of *Drosophila* homeotic genes structurally related to the yeast
transcriptional activator *SNF2/SWI2*. *Cell* 68: 561-572.

5 Thomas, W. K. & Pääbo, S. 1993. DNA sequences from old
tissue remains. *Methods Enzymol.* 224: 406-419.

Tone, M., Nakano, N., Takao, E., Narisawa, S. & Mizuno, S.
1982. Demonstration of W chromosome-specific repetitive DNA sequences
in the domestic fowl *Gallus g. domesticus*. *Chromosoma* 86: 551-69.

10 Troelstra, C., Van Gool, A., de Wit, J., Vermeulen, W.,
Bootsma, D. & Hoeijmakers, J. H. 1992. ERCC6, a member of a subfamily
of putative helicases, is involved in Cockayne's syndrome and preferential
repair of active genes. *Cell* 71: 939-953.

15 Tsuchiya, E., Uno, M., Kiguchi, A., Masouka, K., Kanemori,
Y., Okabe, S. & Mikayawa, T. 1992. The *Saccharomyces cerevisiae* NPS1
gene, a novel CDC gene which encodes a 160kDa nuclear protein involved
in G2 phase control. *Embo J.* 11: 4017-4026.

20

25

30

CLAIMS

- 5 1. The nucleotide sequences of CHD-1A and CHD-W as shown in Fig. 5, Fig. 7 and Fig. 8.
2. A clone or subclones of CHD-1A and CHD-W as defined in 1.
3. A fragment of CHD-1A and CHD-W capable of giving W
10 specific signal on hybridization to a non-ratite bird.
4. A fragment of CHD-1A and CHD-W obtainable by restriction endonuclease digestion thereof and being capable of giving a W specific signal on hybridization to genomic DNA of a non-ratite bird.
5. A clone or subclone of a fragment according to either of
15 claims 3 and 4.
6. A nucleic acid or fragment or oligonucleotide having substantially the sequence of CHD-1A and CHD-W as set out in Fig. 5, Fig. 7 and Fig 8.
7. A clone or a subclone of a nucleic acid or fragment or
20 oligonucleotide according to claim 6.
8. A nucleic acid or fragment or oligonucleotide having substantially the same sequence of the chicken or great tit CHD-gene as set out in Figs 1, 3, 5, 7 or 8.
9. A nucleic acid or fragment or oligonucleotide being capable of
25 giving a W chromosome specific signal on hybridization to the genomic DNA of a non-ratite bird.
10. A nucleic acid or fragment or oligonucleotide according to claim 4 or claim 9 capable of giving W chromosome specific signal on hybridization to the genomic DNA of a chicken, turkey, duck, parrot.

11. A nucleic acid or fragment or oligonucleotide according to any one of claims 4, 9 and 10 capable of giving W chromosome specific signal on hybridization to the genomic DNA of a non-ratite bird under conditions of high stringency.
- 5 12. A nucleic acid or fragment or oligonucleotide according to any one of claims 4, 9 and 10 capable of giving W chromosome specific signal on hybridization to the genomic DNA of a non-ratite bird under conditions of low stringency.
- 10 13. A nucleic acid or fragment or oligonucleotide according to any one of the claims 9 to 13 containing substantially the sequence of the chicken CHD-gene as set out in Fig. 5, Fig. 7 and Fig. 8.
14. A nucleic acid or fragment or oligonucleotide encoding a CHD-protein, fragment thereof or polypeptide containing a CHD-gene or part thereof or encoding a CHD-mimotope protein or fragment thereof or
- 15 CHD-mimotope polypeptide.
15. A process for ascertaining the sex of an embryo, foetus, cell, tissue or organism comprising hybridizing a nucleic acid or fragment or oligonucleotide according to any one of claims 1 to 14 with DNA or RNA of the embryo, foetus, cell, tissue or organism or with cDNA reverse
- 20 transcribed from RNA of the embryo, foetus, cell, tissue or organism or with cDNA or DNA amplified by cloning or polymerase chain reaction from DNA or RNA of the embryo, foetus, cell, tissue or organism.
16. Use of a nucleic acid or fragment or oligonucleotide of any one of claims 1 to 14 in ascertaining the sex of an embryo, foetus, cell,
- 25 tissue or organism.
17. A process for controlling the sex of the progeny of an organism comprising inserting a nucleic acid or fragment or oligonucleotide of any one of claims 1-14 into the genome of the organism or progenitor thereof.

18. Use of a nucleic acid or fragment or oligonucleotide of any one of claims 1 to 14 in controlling the sex of the progeny of an organism.
19. A CHD-protein, fragment thereof or polypeptide containing a CHD-gene of part thereof or a CHD-mimotope protein, fragment thereof or
5 a CHD-mimotope polypeptide.
20. A protein or fragment thereof or polypeptide containing a CHD-chromobox including at least one of the characteristic amino acid residues at position 11, 12, 20, 27 or 31 inside the chromobox or 3, 6, 8, 12-15 or 16 directly downstream of the chromobox when aligned to best
10 effect and as set out in Fig. 11.
21. A protein or fragment thereof or a polypeptide encoded by a nucleic acid or fragment or oligonucleotide according to claims 1-14 and containing a CHD-chromobox
22. A process for controlling the sex of the progeny of an
15 organism comprising supplying exogenously to a cell of the organism or a progenitor of the organism a protein or fragment thereof or a polypeptide according to any one of claims 19-21
23. A process according to claim 22 wherein the protein or fragment thereof or polypeptide is supplied and activates a *CHD-1A* or
20 *CHD-W* target gene.
24. An antibody or fragment thereof against a protein or fragment thereof or polypeptide according to any one of claims 19-21.
25. An antibody producing cell capable of expressing an antibody or fragment thereof according to claim 24.
- 25 26. Use of a protein or fragment thereof or polypeptide according to any one of claims 19-21 or antibody or fragment thereof or cell according to claims 24 or 25 in ascertaining the sex of an embryo cell tissue or organism.

27. A transgenic or chimeric animal having a heterologous nucleic acid or fragment or oligonucleotide according to any one of claims 1 to 14 in the genome of at least the germ cells of the animal.
28. Gametes of an animal according to claim 27.
- 5 29. Progeny of an animal according to claim 27.
30. Progeny according to claim 29 which are transgenic or chimeric and have a heterologous nucleic acid or fragment according to any one of claims 1-14 in the genome of at least the germ cells of the progeny.
- 10 31. A method of controlling the population of a species of bird which comprises introducing an individual member of the species into the population, said individual having a copy or copies of a nucleic acid fragment or oligonucleotide according to any one of claims 1 to 14 integrated on a chromosome (carrier chromosome) be it sex linked or
- 15 autosomal whereby when the male breeds with other individuals of the population the progeny are substantially of one sex or are sexually dysfunctional intersexes.
32. A method according to claim 31 where the nucleic acid integrated into the carrier chromosome is homologous to the native
- 20 *CHD-1A* or *CHD-W* gene of the bird.

Figure 1.

CCCGGTCGGAGGTTTCAAGGAATGACTAGATGTGGCACTTAGTGCCATGGTCTAGTTGAC	60
AAGGTGATGGTTGGTCAAAAGTTGGACTCGATGATCTCAGAGTTTTTTTCCAGCCTTAAT	120
AATTCATATGAATTCTGTAATTTTATTCTTGATCTTTTTGAGCGAAGTTTGTTTGGGGATT	180
TTAGTTTGGTTTCCCTGTCACTGTTTTCTTTCCTTGAACTGACTTTCATTTGCAACATG	240
AGAATTGCTGTATTTGTCAGGTTACAAGTAGTGCAATGGCTGCTTAGAAGTAGTGAGAAA	300
CATTTAGGGAAATACTGGAGTGAAGCAAACACAGTGGTACTGCCAACTGTAGCTTTGGG	360
ATTTGAGGAGCCACAGAGTTGTATATAAATTTGTTTAATGATATCCTGCCCCCTGCCTTCC	420
ATTAATTGCTTGTTTTATGAAACCACTCTTTTTTTTTTTTTTTTTTTTTTTGGCTTCTTCA	480
TATCCTGTGGTAATGAGTTAATGCATTTAGAAGCACATGGCAGAACTAGGAGATCTGTGG	540
ATGACAGTGGTACAGGAGCTCTGAATTTTTTAGATAAACTATGAGAGTGGAACAGAAAT	600
CTGAGGCTAGTTTCTTGAGCTGACTGTAAATTTGTGAGAATATTTTCAAGACTACATTA	660
GTTGTGTGTTTGAGGAAAAATAAAATGTTTAAGTTGTCCATTCCTTGAAACCTCCCGACC	720
GGG	723

2/18

Figure 2.

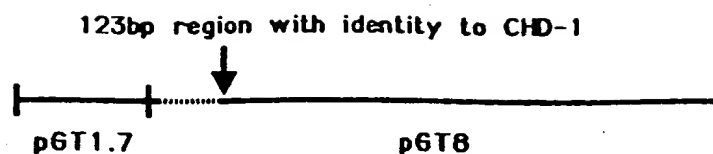


Figure 3.

M	CHD-1	ATTCTTCCAG	ATGATCCTGA	TAAAAAACCA	CAAGCAAAAC	AGTTACAGAC	
C	CHD-1A	ATTTTACCTG	ATGATCCAGA	CAAGAAACCC	CAGGCAAAGC	AGCTACAGAC	
C	CHD-W	ATTTTACCTG	ATGATCCAGA	TAAGAAACCC	CAGGCTAAGC	AGTTACAGAC	
GT	CHD-W	ATTTTACCTG	ATGACCCAGA	TAAGAAACCA	CAGGCAAAGC	AGTTGCAGAC	
M	CHD-1	CAAAAAACCA	CAAGCAAAAC	AGTTACAGAC	CCGTGCAGAC	TACCTCATCA	
C	CHD-1A	CAAGAAACCC	CAGGCAAAGC	AGCTACAGAC	CCGTGCAGAC	TACCTCATTA	
C	CHD-W	CAAGAAACCC	CAGGCTAAGC	AGTTACAGAC	CCGTGCAGAT	TACCTCATTA	
GT	CHD-W	CAAGAAACCA	CAGGCAAAGC	AGTTGCAGAC	CCGTGCAGAT	TACCTCATTA	
M	CHD-1	AACTACTTAG	CAGAGATCTT	GCAAAAAGAG	AGGCTCAGAG	ACTTTGTGGT	GCG
C	CHD-1A	AATTACTGAA	TAAAGACCTT	GCAAGAAAGG	AAGCACAAAG	GCTTGCTGGT	GCA
C	CHD-W	AATTACTGAA	TAAAGACCTT	GCAAGAAAGG	AAGCACAGAG	ACTTGCTGGT	GCA
GT	CHD-W	AATTACTGAA	TAAAGACCTT	GCAAGAAAAG	AAGTGCAAAG	ACTTACTGGT	GCA
M	CHD-1	ILPDDPDKKPQAKQLQTRADYLIKLLSRDLAKREAQRLCGA					
C	CHD-1A	ILPDDPDKKPQAKQLQTRADYLIKLLNKDLARKEAQRLAGA					
C	CHD-W	ILPDDPDKKPQAKQLQTRADYLIKLLNKDLARKEAQRLAGA					
GT	CHD-W	ILPDDPDKKPQAKQLQTRADYLIKLLNKDLARKEVQRLTGA					

Figure 4.

Figure 5.

1 CGGGCTGCGG CACGAAGCGC ACCGCCGGCG CACGCAGGCT CGGGCCGGGG
51 AAGGCCTGGC CCGCCGAGCC GGACGCACGC AGGTATTTGG GCAAAAATCT
101 TGGCCATCTG TAGAGAATAG CAAGTCAAAC GCATTACTTC GAAAACATAC
151 GGAGTACCAG AAAGGGGATT CTTGACCTAC ACCTTGTAAC CTGAGTGGAC
201 TTTCTTTTAA ACTTCTTAAT ACTTACAATG AATGGGCACA GTGATGAAGA
251 AAGTGAAGA AACAGCAGTG GAGAGTCAAG CAGATCAGAT GATGATTCTG
301 GGTCACTTC AGGTTCTGGA TCTGGTTCAA GCTCTGGAAG CAGTAGCGAT
351 GGAAGTAGCA GCCAGTCAGG TAGCAGTGAC TCTGAATCTG GTTCAGAGTC
401 AGGCAGTCAA TCCGAATCAG AGTCTGACAC ATCTAGAGAG AAGAAACAAG
451 TTCAAGCTAA ACCTCCGAAA GCTGACGGAT CTGAGTTTGG GAAGTCCAGT
501 CCAAGCATACT TTGCTGTACA GAGATCAGCA GTGCTCAAGA AGCAACAGCA
551 ACAGCAAAAA GCAGCATCAT CAGACAGTGG TTCAGAAGAG GACTCATCCA
601 GTAGTGAAGA TTCTGCCGAT GATTCTGTCCA GTGAAACTAA GAAGAAAAAG
651 CATAAAGATG AAGACTGGCA AATGTCAGGG TCAGGGTCAG TATCAGGAAC
701 TGGTTCTGAT TCTGAATCGG CGGAAGATGG GGATAAAAAG AGTTGTGAAG
751 AAAGTGAATC TGACTATGAG CCAAAAAACA AAGTCAAAAG CCGTAAACCT
801 CCAAGCAGAA TTAAGCCAAA AAGTGGGAAA AAGAGCACAG GACAGAAGAA
851 GAGGCAACTT GATTCATCAG AGGAGGAGGA GGACGATGAT GAAGATTATG
901 ATAAGAGAGG ATCTCGTCGC CAGGCAACAG TGAATGTTAG TTACAAAGAA
951 GCTGAAGAAA CCAAGACAGA TTCTGATGAT TTGCTGGAAG TTTGTGGAGA
1001 GGATGTCCCA CAGACTGAAG AAGATGAATT TGAAACTATA GAGAAGTTTA
1051 TGGACAGTCG AATTGGCCGA AAAGGAGCCA CTGGTGCCTC AACCACCATC
1101 TATGCCGTG AGGCAGATGG TGACCCAAAT GCTGGGTTTG AAAAGTCAAA
1151 GGAGCTGGGA GAAATACAGT ATCTTATTAA ATGGAAAGGC TGGTCACACA
1201 TCCATAACAC TTGGGAAACT GAAGAAACGC TGAAGCAACA AAATGTTAAA
1251 GGAATGAACA AACTGGACAA CTACAAGAAA AAGGATCAGG AGACAAAACG
1301 CTGGCTGAAA AATGCTTCTC CAGAAGATGT GGAATATTAT AACTGCCAGC
1351 AGGAGCTTAC AGATGATCTG CACAAACAAT ATCAAATAGT GGAAGAATA
1401 ATTGCTCATT CAAATCAAAA GTCAGCAGCT GGTATCCGG ACTACTATTG
1451 CAAATGGCAG GGTCTGCCTT ACTCAGAATG TAGCTGGGAA GATGGTGCTC
1501 TCATTGCCAA AAAGTTTCAG GCACGCATTG ATGAGTATTT TAGCAGAAAT
1551 CAATCCAAGA CTACTCCCTT TAAGGACTGC AAGGTTCTAA AACAGAGACC
1601 AAGATTTGTT GCACTGAAGA AGCAACCATC TTACATTGGA GGACATGAAA
1651 GTCTGGAGTT AAGAGATTAT CAGTTAAATG GATTGAATTG GCTCGCTCAT
1701 TCATGGTGCA AAGGAAATAG TTGTATTCTT GCAGATGAAA TGGGTCTGGG
1751 TAAAACAATA CAAACAATTT CTTTTCTGAA CTACCTGTTT CATGAACATC
1801 AACTGTATGG CCCTTTTCTT CTGCGCGTGC CACTTTCTAC CTTGACATCT
1851 TGGCAAAGAG AGATTCAAAC TTGGGCTCCT CAGATGAATG CTGTAGTTTA
1901 CTTAGGAGAT ATAAGTAGTA GAAATATGAT AAGGACTCAT GAATGGATGC
1951 ATCCACAGAC TAAACGATTA AAGTTTAACA TACTTCTGAC GACATATGAA
2001 ATTTTACTGA AGGATAAGTC ATTCCTTGGT GGTCTCAATT GGGCATTCTAT
2051 AGGAGTTGAT GAAGCTCATC GTTTAAAAAA TGATGACTCT CTTCTGTACA
2101 GGACTTTAAT AGACTTTAAG TCCAACCATC GACTTCTGAT TACTGGAACC
2151 CCACTGCAAA ATTCCCTCAA AGAGCTGTGG TCTTTGTTGC ATTTCTATCAT
2201 GCCAGAAAAA TTTTCTCCTT GGGAAAGATT TGAAGAGGAG CATGGCAAAG
2251 GAAGAGAGTA TGGTTATGCA AGTCTTCACA AAGAGCTTGA ACCATTTTTTA
2301 CTAAGAAGAG TTAAGAAAGA TGAGAAAAG TCTTTACCTG CTAAGGTTGA
2351 ACAAAATCTG AGGATGGAAA TGAGTGCAAT GCAGAAGCAA TATTACAAGT
2401 GGATTTTAAC AAGGAATTAT AAAGCCCTCA GTAAAGGTTT AAAAGGCAGT
2451 ACCTCAGGCT TTCTGAACAT TATGATGGAA CTTAAGAAGT GTTGTAAACCA
2501 TTGCTACCTC ATTAAGCCAC CAGATGATAA TGAATTCTAT AATAAACAGG
2551 AGGCCTTACA GCATTTGATA CGTAGCAGCG GGAAACTAAT CCTTCTTGAC
2601 AAGCTACTGA TTCGTCTGCG AGAACGTGGC AACAGAGTTC TGATTTTCTC
2651 TCAGATGGTG AGGATGCTGG ACATCCTAGC AGAATATCTG AAGTATCGCC
2701 AGTTTCCCTT CCAGAGACTT GATGGATCAA TAAAAAGGGA ATTGAGGAAG
2751 CAAGCACTGG ATCATTTCAA TGCAGAAGGA TCAGAGGATT TCTGTTTTTT
2801 ACTGTCTACA AGAGCTGGAG GATTAGGTAT TAACTTGGCA TCTGCTGACA
2851 CTGTAGTTAT TTTTGATTCT GACTGGAATC CACAGAATGA TCTGCAGGCA
2901 CAGGCGAGAG CTCATAGAAT TGGACAGAAG AAACAGGTTA ATATTTATCG
2951 GCTAGTCACA AAAGGATCAG TAGAAGAAGA TATTCTTGAA AGAGCCAAGA

SUBSTITUTE SHEET (RULE 26)

5/18

```

3001 AGAAGATGGT GCTAGACCAT TTAGTAATTC AGAGAATGGA CACGACAGGA
3051 AAAACTGTTC TGCATACAGG TTCAACTCCA TCAAGCTCTA CACCTTTTAA
3101 TAAAGAAGAG TTATCAGCTA TTTTGAAGTT TGGTGTCTGAG GAACTCTTTA
3151 AAGAACCTGA AGGAGAAGAA CAGGAGCCCC AGGAAATGGA TATAGATGAA
3201 ATCTTGAAGA GAGCTGAAAC TCGGGAAAAT GAGCCAGGTC CATTGACTGT
3251 AGGGGATGAG TTGCTTTCAC AGTTCAAGGT GCGGAACCTT TCCAATATGG
3301 ATGAAGATGA TATTGAGTTG GAACCAGAAA GAAATTCAAG AAATTGGGAA
3351 GAAATCATCC CAGAATCCCA ACGGAGAAGG ATAGAGGAGG AGGAAAGACA
3401 AAAAGAACTT GAAGAAATAT ACATGCTCCC GAGGATGAGA AACTGTGCAA
3451 AACAGATCAG CTTTAATGGG AGTGAAGGAA GACGCAGTAG GAGCAGAAGA
3501 TATTCTGGAT CTGATAGTGA CTCCATCACA GAAAGAAAAA GGCCAAAAAA
3551 GCGTGGAAGA CCTCGAACCA TTCTCGAGA AAATATTAAA GGATTTAGTG
3601 ATGCAGAGAT CAGGCGGTTT ATCAAGAGTT ACAAGAAATT TGGTGGCCCT
3651 CTGGAAAGGT TAGATGCTGT AGCTAGAGAT GCTGAACTGG TTGATAAATC
3701 TGAGACAGAC CTTAGACGTT TGGGTGAACT TGTACATAAT GGATGCATTA
3751 AGGCTTTAAA GGACAATTCA TCTGGACAAG AAAGAGCAGG AGGTAGACTT
3801 GGGAAAGTTA AAGGCCCAAC GTTTCGAATC TCAGGAGTGC AGGTGAATGC
3851 AAAACTAGTC ATCTCTCAGC AAGAAGAGCT GGCACCACTG CACAAATCCA
3901 TTCCTTCAGA TCCAGAAGAA AGGAAAAGAT ATGTCATCCC ATGCCACACC
3951 AAGGCTGCTC ACTTCGATAT AGATTGGGGT AAAGAAGATG ATTCCAATCT
4001 GTTAGTAGGC ATCTATGAAT ATGGCTATGG CAGCTGGGAA ATGATAAAAA
4051 TGGATCCAGA TCTCAGCTTA ACACAGAAGA TTTTACCTGA TGATCCAGAC
4101 AAGAAACCCC AGGCAAAGCA GCTACAGACC CGTGCAGACT ACCTCATTA
4151 ATTACTGAAT AAAGACCTTG CAAGAAAGGA AGCACAAAGG CTTGCTGGTG
4201 CAGGCAATTC CAAGAGAAGG AAGACAAGAA ATAAGAAGAA TAAGATGAAG
4251 GCTTCAAAAA TAAAAGAAGA AATAAAGAGT GATTCTTCAC CACAACCCTC
4301 AGAAAAATCT GATGAAGATG ATGAGGAGGA GGATAACAAG GTAAATGAAA

```

**

```

4351 TCAATCTGA AAATAAGAA AAATCTAAAA AAATTCATT GCTGGATACT
4401 CCAGTTCATA TTACTGCAAC CAGTGAACCA GTTCCTATCT CAGAAGAATC
4451 TGAAGAATC CATCAGAAGA CATTTAGTGT GTGCAAAGAA AGAATGAGGC
4501 CTGTCAAAGC AGCACTGAAA CAGCTGGATA GACCAGAGAA GGGCCTTTCT
4551 GAAAGGGAGC AGCTGGAACA TACTAGGCAG TGCTAATCA AAATTGGGGA
4601 TCACATTACA GAATGCCTGA AGGAGTACAC AAATCCCAG CAAATAAAAC
4651 AGTGGAGGAA AAATTTGTGG ATTTTGTGT CCAAGTTTAC AGAATTTGAT
4701 GCCAGAAAGC TGCACAAACT CTACAAACAT GCAATCAAAA AGCGCCAAGA
4751 GTCTCAGCAA CACAATGACC AAAACATTAG CAGCAATGTG AATCACATG
4801 TAATCAGAAA TCCAGATGTG GAAAGACTGA AGGAGACTAC AAACCATGAT
4851 GATAGTAGCA GGGACAGTTA TTCTTCTGAT AGACATTTAT CACAATACCA
4901 TGATCATCAC AAAGACAGGC ATCAGGGAGA TGCTTACAAG AAAAGTGACT
4951 CCAGGAAAAG GCCATATTCA GCCTTCAGTA ATGGAAAAGA TCACAGAGAC
5001 TGGGATCACT ACAAACAGGA CAGCAGATAC TACAGTGATA GTAAACATAG
5051 AAAGTTAGAT GACCACAGGA GCAGAGACCA CAGGTCAAAC CTGGAAGGAA
5101 ACTTAAAGA CAGCCGGGGT CATTGAGATC ACCGCTCCCA TTCAGACCAC
5151 AGGATACACT CAGATCACCG TTCCACTTCA GAATACAGCC ATCATAAATC
5201 TTCGAGAGAT TATAGATACC ACTCAGACTG GCAAATGGAC CACAGAGCTT
5251 CTGGTAGTGG CCCGAGGTCA CCCTAGATC AGAGGTCTCC TTATGGTTCA
5301 AGATCTCCCC TAGGACACAG ATCTCCATTT GAACACTCAT CAGATCACAA
5351 AAGTACACCT GAACATACAT GGAGTAGCCG GAAGACATAA CAAAGACTGA
5401 CATTTTCTGG ACCTTCTTTT TAGCCATATA CAGTAAACTA ACACAGTAAT
5451 TGCCTTACAT GACTTGAAG ATATGGACTG GATATTCTAT CAGTAGCAGT
5501 ATTGTTACTT CTTTCCAGGA TGCAAGGTCT ATTATCCCAA CAGAAGAAAA
5551 ATATTTTTGT ATTTAAAGTT TATGCTGCAC TGTGCTGCAA ATGTTGTGGC
5601 ACTTTTTTTT TAAGAAATGG AAGATGTTTA CTTTACAGG GACCTCAACA
5651 CTGCCCCTTT CAGACTGGAT CTTACTATAA AACTCTTCAT GTCAAAGTGG
5701 TTCTAGGCTG AACACAGATT AAATTATGTT TGTAATGAA CACTTAAACA
5751 CTGACCTGTG CTTATGTTT AGGAAAGAAT GGGGATTTA TTTTGTTTA
5801 TTTCTTGGTA GAGAACTCTC AAGGACTTTG TTCACTTTCC AAAGCTACTT
5851 GTTTACATTG TACACTGCGA CCACCTTGCC GCTTTTCATC ACAAGCTTGA
5901 ATATTTAAAT TCTGTACCTA CAGTTGTAAA ATAGCCAGGA TTTCTCCTGT
5951 TTGTGATCAG TTATAATGCC TTTTATGAA ACAAACAAAC AAACAAAAAA
6001 CAATTAATAA AAAAAACACA ACAAACCAA CAAATGGCTG TAAATTATTG
6051 TAAATTAATT AAATGAGCTT TTTTCCGTCA GGCTTTTTTT GGCTGTTTCT
6101 TTCCCCAACA ACTCAGGCCT TCTTTTCACA AAGTCAGTAT ACTTACATGT
6151 TTTAATAAAA TATCTCGATG GAATCAGAAT GTAAAAATGG GGAAGGGAAT
6201 ATTTTATTCC ATTTAGTGCT CTTTTTTTAT TGGATACTTT TACATACCTG
6251 TTTTGGTTG TTTTATTTTA TTTTTTTTTT CTATTAAACT GTCAGTGTG
6301 TGATTGTTGT AATGAACAGT GAGAATATCC CACTCTAAAC GTGCCCTGG
6351 AAAGCTTTTC AGGTGCATTG GTTTAAAGA AGGAAGTGT CTATAGGTGA

```

SUBSTITUTE SHEET (RULE 26)

6/18

6401 ACACTTCAAA ACCCAGATCA GCCAAGATTC ATTGTAAATC CATTTGTTTT
6451 CCCTCTTTAA CATGGGCAAT AATGTCAAAT GTGCTATGCA GCAGTTAATA
6501 TTTTAGAAGA TTTGAATGAC TTTATTAACA GAATTGTTAC AATGCACACT
6551 GATTGTACAT AGATAACTTC TATCTGACAA ATTAAATTAA CTAAAACCAA
6601 AAAAAACC

7/18

Figure 6.

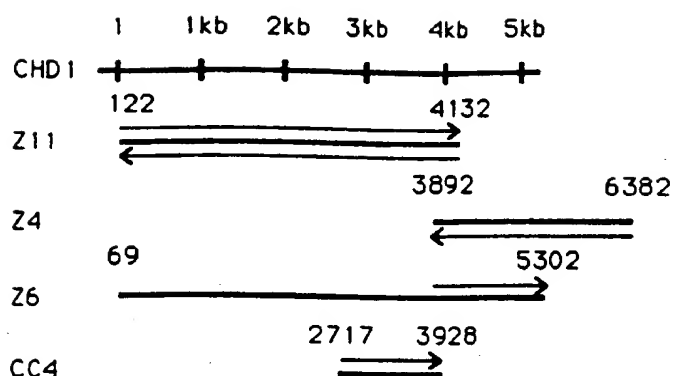


Figure 7.

		D E I V S V K H L H K K I K T E
CHD-1A	1	GATGAGATTGTTTCAGTGAAACATCTACATAAAAAAATAAAAACAGAAA
CHD-W	1	GATGGGATTGTTTCAGTGAAACATCCACATAAAAAAATAAAAGCAGAAA
		D G I V S V K H P H K K I K A E
		K E N E E K P E P D I G I K K E A
CHD-1A	51	AAAGAAAATGAAGAAAAGCCTGAGCCAGATATTGGTATAAAGAAGGAAGCT
CHD-W	51	AAAGAAAATGAAGAAAAGATGAGCCAGAGATTGGTATAAAGAAGGAAGCT
		K E N E E K D E P E I G I K K E A
		E E K R E T K E K E N K R E L K R
CHD-1A	101	GAAGAAAAAGAGAGACAAAAGAGAAGGAAAATAAAAGGGAATTGAAAAGG
CHD-W	101	GGAGAAAAAGAGAGACAAAAGAAAAGGAAAATAAGA
		G E K R E T K E K E N K
		E K K E K E D K K E L K E K D N K
CHD-1A	151	GAGAAAAAGAAAAGAGGATAAGAAAGAATTAAAAGAAAAGATAATAAA
		E K R E N K V K E S T Q K E K E V
CHD-1A	201	GAAAAGAGAGAAAACAAAGTAAAGAATCCACACAGAAAGAAAAGGAAGTG
		K E E K
CHD-1A	251	AAGGAAGAGAAG

Figure 8.

ATTTATCGGC	TAGTCACAAA	AGGATCAGTA	GAAGAAGATA	TTCTTGAAAG	AGCCAAGAAA	AAGATGGTGT	TAGATCATT
10	20	30	40	50	60	70	80
AGTGATTGAG	AGAATGGACA	CCACAGGGAA	AACTGTACTA	CATACAGGCT	CTACTCCTTC	AAGCTCAACA	CCTTTTAATA
90	100	110	120	130	140	150	160
AGGAAGAGTT	ATCAGCAATT	TTGAAGTTTG	GTGCTGAGGA	ACTTTTAA	GAACCTGAAN	NNGAAGAAGA	GGAGCCTCAG
170	180	190	200	210	220	230	240
GAGATGGATA	TAGATGAAAT	OCTGAAGAGG	NCTGAACTC	GAGAAAATGA	GTCAGGCCCA	TTAACTGTAG	GAGATGAGTT
250	260	270	280	290	300	310	320
ACTTTCACAG	TTCAAGGTAG	CTAACTTTTC	CAATATGGAT	GAAGATGACA	TTGAATTGGA	ACCAGAACAA	AATCTAAGAA
330	340	350	360	370	380	390	400
ACTGGGAAGA	AATCATTCCA	GAAGTTCAGT	GGCAGCGAAT	AGAGGGGNG	GAAAGACAAA	AAGAACTTGA	AGAAATATAT
410	420	430	440	450	460	470	480
ATGCTTCCAA	GAATGAGAAA	CTGTGCAAAA	CAGATCAGCT	TTAATGGAAA	TGAAGGGAGA	TGCAGTAGGA	GCAGAAGATA
490	500	510	520	530	540	550	560
TTCTGGATCT	GATAGTGATT	CCATCTCAGA	AAGAAAACGA	CCAAAAAAC	GTGGACGACC	ACGAATATT	CCCCGTGAAA
570	580	590	600	610	620	630	640
ACATTAAAGG	ATTTAGTGAT	GCAGAGATTA	GACGATTAT	CAAGAGTTAC	AAGAAATTTG	GTGGCCCGAT	TGAAAGGTTA
650	660	670	680	690	700	710	720
GATGCTATAG	CTAGAGATGC	TGAGCTAGTT	GATAAATCTG	AAACAGACCT	TAGACGCTG	GGAGAATTG	TACATAATGG
730	740	750	760	770	780	790	800
ATGCATTAAG	GCTTTAAATG	ATAATGACTT	TGGTCAAGGA	AGAACAGGTG	GTAGATTTCG	GAAAGTTAAA	GGCCCAACAT
810	820	830	840	850	860	870	880
TCCGAATAGC	AGGAGTGCAG	GTGAATGCAA	AGCTAGTCAT	TTCTCAGGAA	GAAGAGTTGG	CACCATGCA	TAAATCGATT
890	900	910	920	930	940	950	960
CCTTCAGATC	CAGAAGAAG	GAAAAGATAT	GTCATCCCAT	ACCACACCAA	AGCAGCTCAT	TTTGATATAG	ATTGGGGTAA
970	980	990	1000	1010	1020	1030	1040
AGAAGATGAT	TCCAATCTGT	TAATAGGCAT	CTATGAATAT	GGTTATGGCA	GTTGGGAAAT	GATAAAATG	GATCCTGATC
1050	1060	1070	1080	1090	1100	1110	1120
TCAGTTTGAC	ACAGAAGATT	TIACCTGATG	ATCCAGATAA	GAAACCCAG	GCTAAGCAGT	TACAGACTOG	TGCAGATTAC
1130	1140	1150	1160	1170	1180	1190	1200
CTCATTAAT	TACTGAATAA	AGACCTTGCA	AGAAAGGAAG	CACAGAGACT	TGCTGGTGCA	GGCAATCAA	AGAGGAGAAA
1210	1220	1230	1240	1250	1260	1270	1280
AACAAGAAGT	AAGAAGAATA	AAGCAACAAA	GGCTGC				
1290	1300	1310					

Figure 9.

[illegible]

10/18

C CHD-1A
M CHD-1

DBFNAEGSEDFCFLLSSTRAGGLGINLASADTVVIFDSWNPQNDLQAQARAERIQQKQV
DBFNAEGSEDFCFLLSSTRAGGLGINLASADTVVIFDSWNPQNDLQAQARAERIQQKQV

C CHD-W
C CHD-1A
M CHD-1

-IYRLVTGGSVEEDILERAQKKMVLDBLVIQRMOTTGKTVLHTGSTPSSSTPFNKEELSA
NIYRLVTGGSVEEDILERAQKKMVLDBLVIQRMOTTGKTVLHTGSAPSSSTPFNKEELSA
NIYRLVTGGSVEEDILERAQKKMVLDBLVIQRMOTTGKTVLHTGSTPSSSTPFNKEELSA

C CHD-W
C CHD-1A
M CHD-1

ILKFGAEELFKPEPEXEEEPQEMDIDEILKRAETRENESGPLTVGDELLSQFKVANFSNM
ILKFGAEELFKPEPEGEZEPQEMDIDEILKRAETRENESGPLTVGDELLSQFKVANFSNM
ILKFGAEELFKPEPEGEZEPQEMDIDEILKRAETRENESGPLTVGDELLSQFKVANFSNM

C CHD-W
C CHD-1A
M CHD-1

DEDDIELEPEQNLNRWEEIIPVQWRRIEGKERQKELEEIYMLPRMRNCARDISFNGSEG
DEDDIELEPERNSKNWEEIIPVQWRRIEGKERQKELEEIYMLPRMRNCARDISFNGSEG
DEDDIELEPERNSKNWEEIIPVQWRRIEGKERQKELEEIYMLPRMRNCARDISFNGSEG

C CHD-W
C CHD-1A
M CHD-1

RCSRSRRYSGSDSDSISERKRPKRGPRPTIPRENIGFSDAEIRRFIKSYKFGGPVER
RCSRSRRYSGSDSDSISERKRPKRGPRPTIPRENIGFSDAEIRRFIKSYKFGGPVER
RCSRSRRYSGSDSDSISERKRPKRGPRPTIPRENIGFSDAEIRRFIKSYKFGGPVER

C CHD-W
C CHD-1A
M CHD-1

LDALARDAEVLVDKSETDLRRLGELVNGC IKALNDNDFGQGRGGRFGVKGPTRFRIAGV
LDALARDAEVLVDKSETDLRRLGELVNGC IKALNDNDFGQGRGGRFGVKGPTRFRIAGV
LDALARDAEVLVDKSETDLRRLGELVNGC IKALNDNDFGQGRGGRFGVKGPTRFRIAGV

C CHD-W
C CHD-1A
M CHD-1

QVNAKLVISHEEELAPLEHSIPSDPEERKRYVIPYHTAAHFDIDWGEKDDSNLLIGIYE
QVNAKLVIHAEDELIPLEHSIPSDPEERKRYVIPYHTAAHFDIDWGEKDDSNLLIGIYE
QVNAKLVISHEEELAPLEHSIPSDPEERKRYVIPYHTAAHFDIDWGEKDDSNLLIGIYE

C CHD-W
C CHD-1A
M CHD-1

YGYGSWEIMKMDPDLSTOKILPDDPDKKPQAKQLQTRADYLIKLLNKDLARKEAQRLAG
YGYGSWEIMKMDPDLSTOKILPDDPDKKPQAKQLQTRADYLIKLLNKDLARKEAQRLAG
YGYGSWEIMKMDPDLSTOKILPDDPDKKPQAKQLQTRADYLIKLLNKDLARKEAQRLAG

C CHD-W
C CHD-1A
M CHD-1

AGNSKRRKTRSKKKAATAA
AGNSKRRKTRAKSKAMKS IKVKEEIKSDSSPLPSEKSDDEDD---KNDSEKESKDRS
AGNSKRRKTRAKSKAMKS IKVKEEIKSDSSPLPSEKSDDEDD---KNDSEKESKDRS

C CHD-1A
M CHD-1

KKSVDAPVHITASGEVPVIAEZEELDQRTTSICKERMRPVKAALKQDRPEKGLSER
KKSVDAPVHITASGEVPVIAEZEELDQRTTSICKERMRPVKAALKQDRPEKGLSER
KKSVDAPVHITASGEVPVIAEZEELDQRTTSICKERMRPVKAALKQDRPEKGLSER

C CHD-1A
M CHD-1

EQLEHTROCLIKIGDHITECLKEYSNPEQIKQWRJNLWIFVSKFTEFDARKLEKLYKHA
EQLEHTROCLIKIGDHITECLKEYSNPEQIKQWRJNLWIFVSKFTEFDARKLEKLYKHA
EQLEHTROCLIKIGDHITECLKEYSNPEQIKQWRJNLWIFVSKFTEFDARKLEKLYKHA

C CHD-1A
M CHD-1

KKRQESQONSQON-SNVATTHVIRNPDMERLKENINHDSSRDSYSSDRHLSQYHDHKKD
KKRQESQONSQON-SNVATTHVIRNPDMERLKENINHDSSRDSYSSDRHLSQYHDHKKD
KKRQESQONSQON-SNVATTHVIRNPDMERLKENINHDSSRDSYSSDRHLSQYHDHKKD

C CHD-1A
M CHD-1

RHQGDSYKKSRSRKPYSFSGKDERWDHYRQDSRYSDREKRRKLDHRSREHRPSL
RHQGDAYKKSRSRKPYSFSGKDERWDHYRQDSRYSDREKRRKLDHRSREHRPSL
RHQGDAYKKSRSRKPYSFSGKDERWDHYRQDSRYSDREKRRKLDHRSREHRPSL

C CHD-1A
M CHD-1

EGGLKD-RCHSDHRSBHDHMSDHRSTPSTHINPPRDYRYSWQDERAASSGPRSP
EGNLKDSRGHSDHRSBHDHMSDHRSTPSTHINPPRDYRYSWQDERAASSGPRSP
EGNLKDSRGHSDHRSBHDHMSDHRSTPSTHINPPRDYRYSWQDERAASSGPRSP

C CHD-1A
M CHD-1

LDQSPYGRSP-----FEHSAEHRSTPEHTWSSRKTQKLMSSGTLFXP
LDQSPYGRSPPLGHRSPFEHSDHSTPEHTWSSRKTQKLMSSGTLFXP
LDQSPYGRSPPLGHRSPFEHSDHSTPEHTWSSRKTQKLMSSGTLFXP

C CHD-1A
C CHD-1A
C CHD-1A
C CHD-1A
C CHD-1A
C CHD-1A
C CHD-1A

LTXLERYGLDILSVAVILLLSRMQGLLSQOKKNI FVYKVYAALCCKCCGTFILNRCGLL
LQGPQHCFFQTSYKTLVKKVVLGXTOIKLCLXMTXLTCAVYSGRNGGFILFYFLVE
NSQGLCSLSKATCLBCTLRPPCFRSSQAXIFKFCYSCKLARI SPVCDQLXCLFMKQTNK
QKTIKQNTTKPTNQCKLLXINXMSFPFSGFWLFLSPTQAFPSQSYTYNFXNISM
SECKNGEKNILFBLVLLFYWILBCTFWLFFYIFFFYXTVSVVIVVHNSENIPLXTVPWK
AFQVHWYKRRKCSIGEHFKTKIQSDSLXIELFLSLFMGNVVKCAMQQLIFXKIXMTLLTE
LLQCTLIVERXLLSDKLNXLKPKT

SUBSTITUTE SHEET (RULE 26)

```

55 KPPKADGSEFWKSSPSILAVQRSVAVLKQOQQQKAASSDGSSEEDSSSE 104
.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:
2654 MAAKDISTEVLQN.PELYGLRRS...HRAAAHQQNYFNDSDDDD...E 2695
105 DSADDSSSETKRRKKHDEDDWQMSGSGSVSGTGSDESSEAEDGDKSSCEESE 154
|.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.
2696 DNIKQSRKRMTTIEDDED.....EFEDEEGEEDSGEDEDEEDFEEDD 2738
155 SDYEPKNKVKSRKPPSRIPKPSGKSTGQKKRQLDSSEEEEDDDDEDYDKR 204
.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:
2739 DYYGSPKQNRSKPKSRTKSKSKSKPKSQSEKQSTVKIP.....TRF 2780
205 GSRRQATVNVSYKEAEETKTDSDDLLE...VCGEDVPQT...EEDEFE 246
.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:
2781 SNRQNKTVNYNIDYSDDDLLESEDDYGSEELSEENVBEASANPQPEDFH 2830
247 TIEKFMDSRIGRKATGASTTIYAVEADGDPNAGFEKSKELGEIQYLIKW 296
|.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.
2831 GIDIVINBRL.....KTSLEEGKVLEKTVPDNNCKE..NYEFLIKW 2870
297 KGWSHIHNTWETEETLKQNVKGMNKLDNYYKK...DOETKRWLKNASPE 343
.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:
2871 TDESHLHNTWETYESIGQ..VRGLKRLDNYSKQFIIEDQOVRDPYVTAE 2918
344 DVEYYNCQOELTDDLHKQYQIVERIIA..HSNQKSAAGYPDYCKWQGLP 391
|.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.
2919 DIEIMDMERERRLDEFEEFHVPERIIDSQRASLEDGTSLOQLVVKWRRLN 2968
392 YSECSWEDGALIAKKFOARIDEYPSRNQSKTTPFKDCKVLKQRPFRVALK 441
|.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.
2969 YDEATWENATDIVKLAPEQVKHFQNRRENSKILPQYSSNYTSQRPRFEKLS 3018
442 KQPSYIGGHESELELRDYQLNGLNWLHBSWCKGNSCILADEMGLGKTIQTI 491
|.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.
3019 VQPPFIKGG...ELRDFQLTGINWMAFLWSKGDNGILADEMGLGKTVQTV 3065
492 SFLNYLFHEHQLYGPFLLRVPLSTLTSTWQREIQTWAPQMNNAVVLGDITS 541
|.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.
3066 AFISWLIFARRONGPHIIVVPLSTMPAWLDTFEKWAPDLNCICYMGNQKS 3115
542 RNMIRTHEW...MHPQTKRLKFNILLTTYEILLKDKSFGLGLNWAFIGV 587
|.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.
3116 RDTIREYEFYTNPRAGKKTMTKFNVLTTTYEYILKDRaelGSIKWQFMAV 3165
588 DEABRLKNDDSLlyRTLIDFKSNHRLITGTPLQNSLkELWSLLHFIMPE 637
|||||.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:
3166 DEABRLKNAESSLYESLNSPKVANRMLITGTPLQNNIKELAAVLNVLMPG 3215
638 KFSSWEDFE.EEHGKGREYGYASLHKELEPFLLRRVKKDVEKSLPAKVEQ 686
|.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.
3216 RFTIDQEI DFENQDEEQEYIHDLHRRIQPFILRLKKDVEKSLPSKTER 3265
687 ILRMEMSALQKQYKWLTRNYKALSKGSKGSTSGFLNIMMELKKCCNEC 736
|||||.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:
3266 ILRVELSDVQTEYYKNILTKNYSALTAGAKGGHFSLLNIMNELKKASNEP 3315
737 YLIKPPDDNEF.....YNKQeALQHLIRSSGKLILLDKLLIRLRERN 779
|.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.
3316 YLFdNAEERVLQKFGDGKMTRENVRLGLIMSSGKMVLLDOLLTRLLKKGH 3365

```

780	RVLIFSOMVRMLDILAEYLKRYQFPFQRLDGSIKGELRKQALDHFNAEGS	829
3366	RVLIFSOMVRMLDILGDYLSIKGINFQRLDGTVPSPAQRISIDHFNSPDS	3415
830	EDFCFLLSTRAGGLGINLASADTVVIFDSDWNPONDLOAQAARAHRIQOKK	879
3416	NDFVFLSTRAGGLGINLMTADTVVIFDSDWNPQADLOAMARAARIGQKN	3465
880	QVNIYRLVTKGSVEEDILERAKKKMVLDBLVIQRMDDTTGKTVLHTGSTPS	929
3466	HVMVYRLVSKDTEVEEVLERARKKMILEYAIISLGVTDGNKYTKKNEP..	3513
930	SSTPFNKKEELSAILKFGAEELFKPEGEEQEPQEMDIDEILKRAETRENE	979
3514NAGELSAILKFGAGNMFTATD.NQKKLEDNLDDVLNBAEDHVT	3557
980	PG...PLTVGDELLSQFKVANFSNMDEDDIELEPERNSRNWEEIIPESQR	1026
3558	PDLGESHLGGEEFLKQFEVTDY.....KADIDWDDIIPHEEL	3594
1027	RRIEEEERQKELEE.....IYMLPRMRNCAKQI..SFNGSE.....	1060
3595	KKLQDEEQKRKDEEYVKEQLEMMNRDNLAKKIKNSVNGDGTAA NSDSD	3644
1061	..GRRSRRRYSGSDSDSITERKRPKRGRPRPTIPR.ENIKGFSD...AE	1104
3645	DSTSRSSRRRARANDMSIGE...SEVRALYKAILKFGNLKEILDELIAD	3691
1105	IRRFIKSYKKFGGPLERLDAVARDA.....ELVDKSETDLRRLGEL	1145
3692	GTLPVKSFEKYGETYDEMMEAARDCVHEEEKNRKEILEKLEKHATAYRAK	3741
1146	VHNGCIKALKD.NSSGQERAGGRLGKVKGPTFRISGVQ.VNAKLVISHEE	1193
3742	LKSGEIKAENQPKDNPLTRLSLKKREKRAVLNFKGVKSLNAESLLSRVE	3791
1194	ELAPLHKSIPSD.PEERKRYVIPCHTKAA..HFDIDWGKEDDSNLLVGIY	1240
3792	DLKYLKLNLSINYKDDPLKFSLGNNTPKPVQNWSSNWTKEEDEKLLIGVF	3841
1241	EYGYGSWEMIKMDPDLSLTQKILPDD.....	1266
3842	KYGYGSWTQIRDDPFLGITDKIFLNEVENPVAKKSASSSDTTPTPSKKGK	3891
1267PDKKPQAKQLQTRADYLIKLLNKDLARK.....EAQRLAGAGNS	1305
3892	GITGSSKKVPGAIBLGRVDYLLSFLRGGLNTKSPSADIGSKKLPTGPSK	3941
1306	KRRKTRNKKNKMKASKIKEEIKSDSSPQPSEKSDDEDEE...EDNKVNEM	1352
3942	KRQRKPANESKSMTPETISSEPANGPPSKRMKALPKGPAALINNTRLSPN	3991
1353	KSENKEKSKKIPLLDTPVHITATSEVPVPISEESEELBQKTFVUCKERM RP	1402
3992	SPTPPLKSKVSRDNGTR....QSSNPSSGSAHEKEYDSMDEEDCRBTMSA	4037
1403	VKAALKQLDRPEKGLSEREQLEHTRQCLIKIGDHITECLKEYTNPEQIKQ	1452
4038	IRTSLKRLRRGGKSLDRKEWAKILKTELTTIGNBI.ESQKGSSRKASPEK	4086
1453	WRKNLWIFVSKF..TEFDARKLHKLKHAIKKRQESQ	1488
4087	YRKHLWSYSANFWPADVKSTKLMAMY....DKITESQK	4120

13/18

Figure 11.

CCHD	AVEAD	GDPNAGFEKSKELGE.IQYLIKWKGWSHIENTWETEET	LKQQNVKGMNKLDNYKK
MCHD	AVEAD	GDPNAGFERNKEPGD.IQYLIKWKGWSHIENTWETEET	LKQQNVKGNKKLDNYKK
YCHD	EGKVL	EKTVPDLNNCKE..N.YEFLIKWTDESHLENTWETYES	IGQ..VRGLKRLDNYCK
		** **** ** ***** *	* * * ***** *
DHP1	EEEE	YAVEKIIDRRVRK GK.VEYLLKWKGYPETENTWEPENN	LDCQDLIQY
HHP1	EDEE	YVVEKVLD RR VVK GKQVEYLLKWKGFSEEHNTWEPEKN	LDCPELISEF
MMOD1	EEEE	YVVEKVLD RR VVK GK.VEYLLKWKGFSDONTWEPENN	LDCPDLIAEF
MMOD2	AEPE	FVVEKVLD RR VVNGK.VEYFLKWKGF TDADNTWEPENN	LDCPELIEDF
	**	*** **** ** ***** ***** *	*** **
DPC	PVDLV	YAAEKIIQKRVK GK.VEYRVKWKGNQRYNTWEPENN	ILDRRLIDIY
MMOD3	VGEQV	FAAECILSKRLR K GK.LEYLVKWRGWSSKHNSWEPEEN	ILDPRLLLAF
	*	*** * ** *** ** ***** * ***** *	*****
		** ** ** **	

Figure 12.

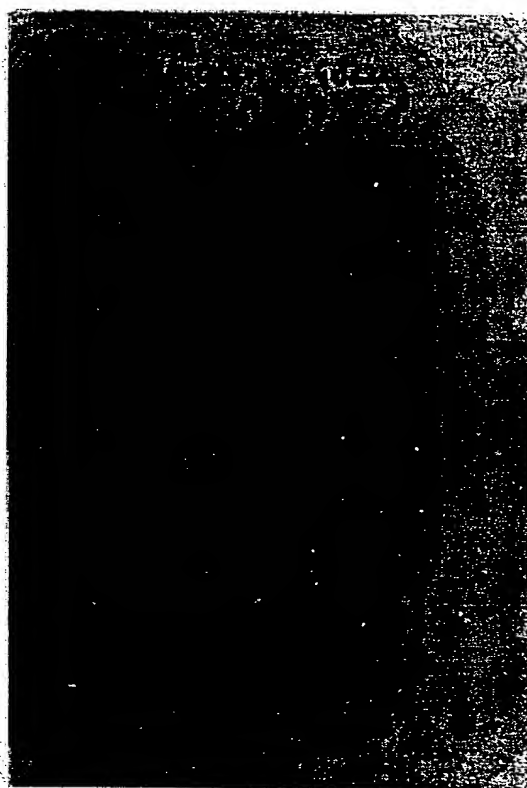
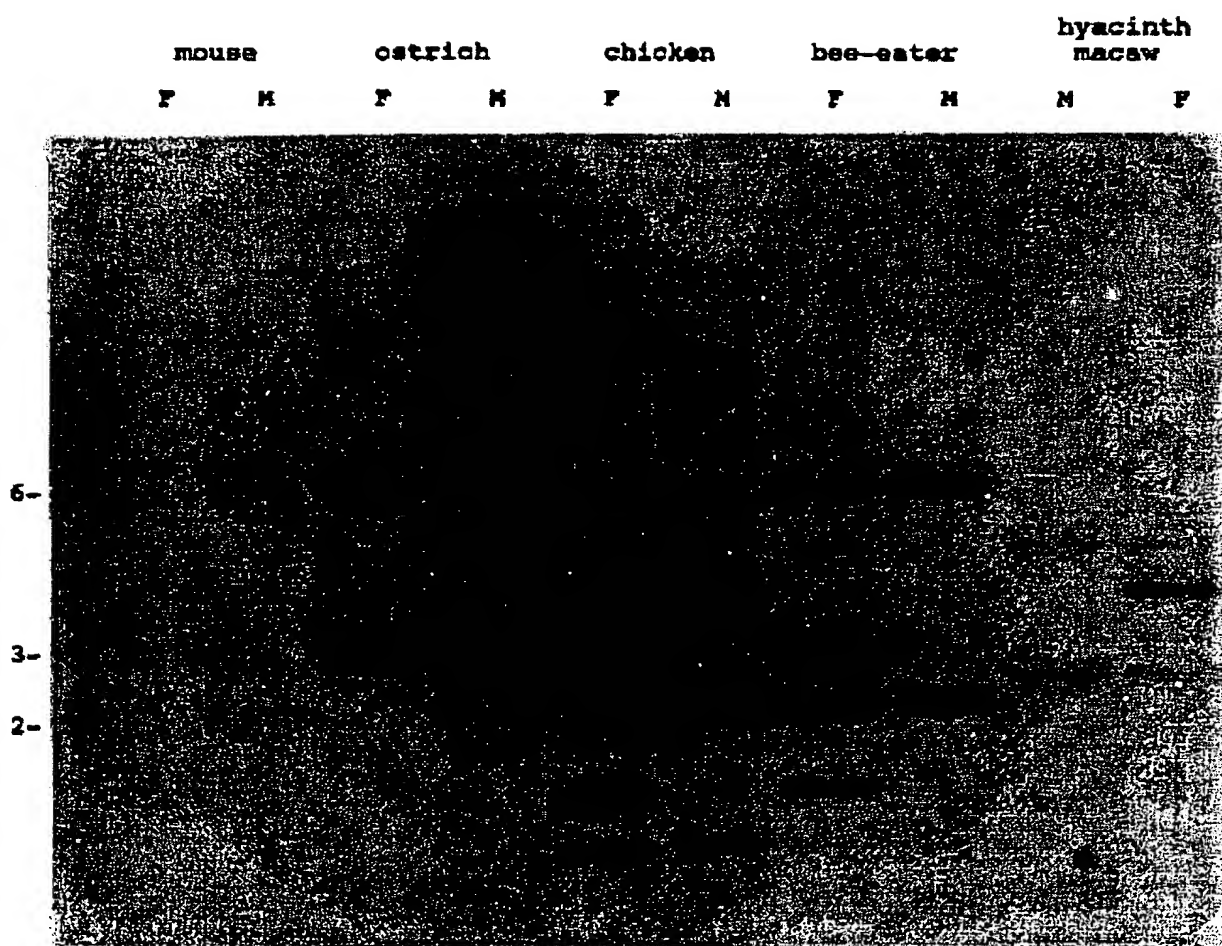


Figure 13.



16/18

Figure 14.

MOUSE	CHD1	AGA TAT TCT GGA TCT GAT AGT GAT TCA ATC TCG GAA
CHICKEN	CHD-1A	--- --- --- --- --- --- --- --C --C --- A-A ---
SPIX	CHD-1A	--- --- --- --- --- --- --- C --C --- --- ---
CHICKEN	CHD-W	--- --- --- --- --- --- --- --C --- <u>---A---</u>
SPIX	CHD-W	--- --- --- --- --- --- --- C --C --- <u>---A---</u>
HYACINTH	CHD-W	--- --- --- --- --- --- --- --C --C --- <u>---A---</u>
P1		A TAT TCT GGA TCT GAT AGT GAT TC
P3		AGA TAT TCC GGA TCT GAT AGT GA
MOUSE	CHD1	AGG AAA CGG CCG AAG AAA CGT GGG CGA CCC CGC ACT
CHICKEN	CHD-1A	--A --- --- <u>---A</u> --A --G --- --A A-- --T --A --C
SPIX	CHD-1A	--- --- --- <u>---A</u> --A --G --- --A A-- --A --A ---
CHICKEN	CHD-W	--A --- --A --A --A --- --- --A --- --A --A ---
SPIX	CHD-W	--A --- --A --A --A -GA --- --- --A --- --A --A ---
HYACINTH	CHD-W	--A --- --A --A -GA --- --- --A --- --A --A ---
MOUSE	CHD1	ATC CCT CGG GAG AAT ATT AAA GGA TTT AGT GAT GCG GAG
CHICKEN	CHD-1A	--T --- --A --A --- --- --- --- --- --- --A ---
SPIX	CHD-1A	--T --- --A --A --- --A --- --- --- --- --- --A ---
CHICKEN	CHD-W	--T --C --T --A --C --- --- --- --- --- --A ---
SPIX	CHD-W	--T --- --T --A --- --- --- --- --- --- --A ---
P2		TTT CCT AAA TCG CTA CGT CT
HYACINTH	CHD-W	--- --- --- --- --- --- --- --C --- --A --G
HYACINTH	CHD-W	ATT AGG CGG T

17/18

Figure 15.

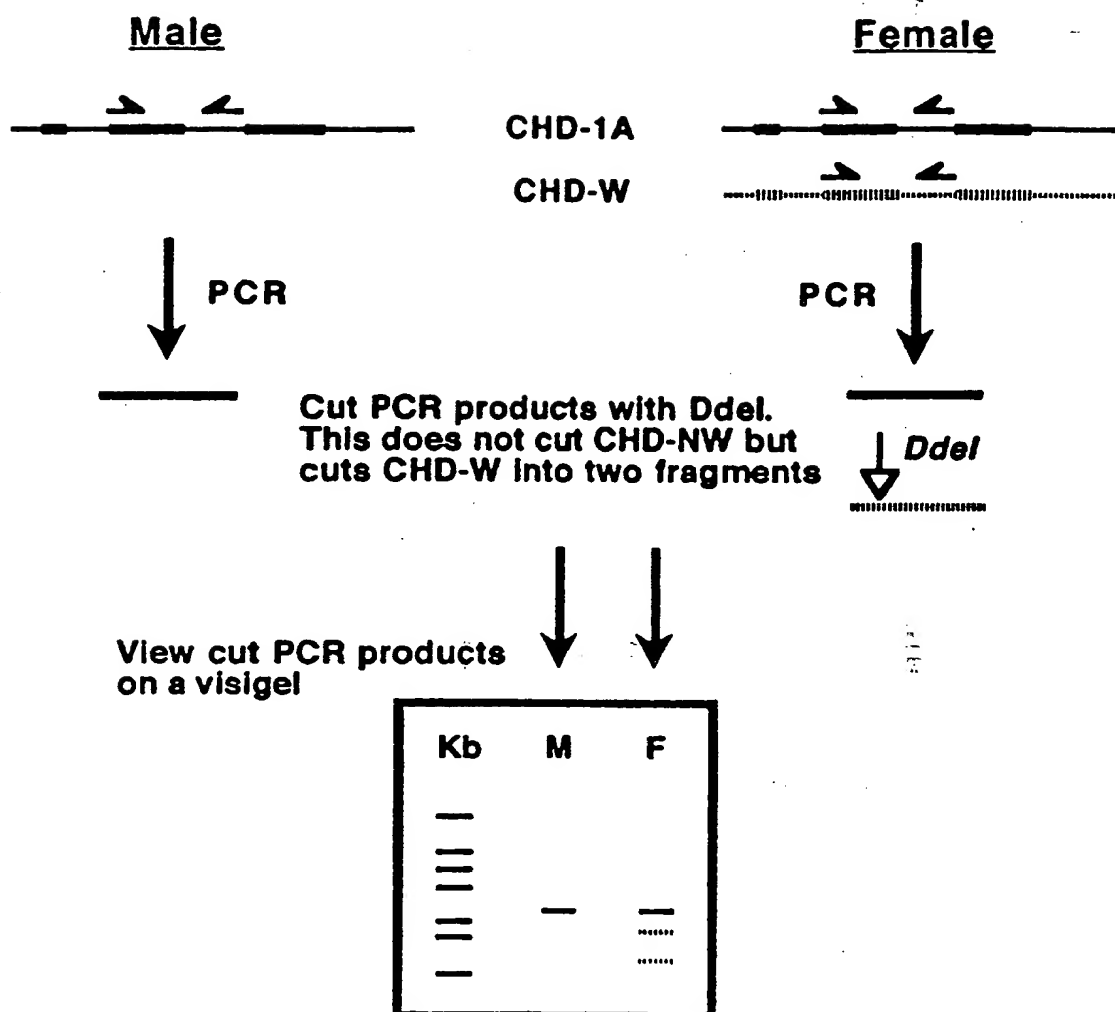


Figure 16.

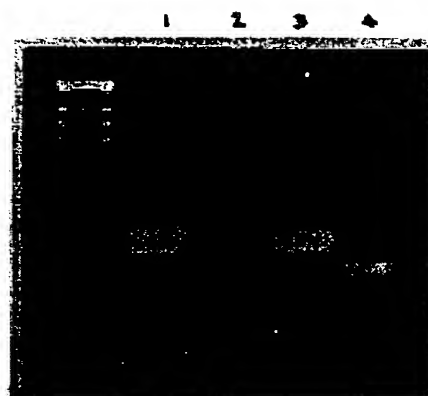
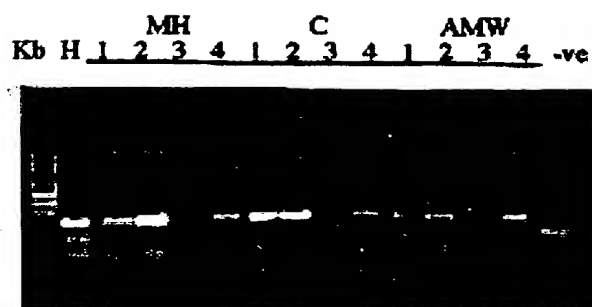


Figure 17.



INTERNATIONAL SEARCH REPORT

International Application No
PCT/GB 96/01341

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/12 C12Q1/68
C07K16/18

C07K14/465 C07K14/47 A01K67/027

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12Q C07K C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EMBL DATABASE, Accessionnumber D14316 Sequence reference GGJF11 from G. gallus 6 August 1993. XP002016259 compare nucleotides 5-2292 with nucleotides 2221-4508 in figure 5. --- -/--	1-14,19

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

* Special categories of cited documents :

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document but published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed

- *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- * & * document member of the same patent family

Date of the actual completion of the international search

17 October 1996

Date of mailing of the international search report

30.10.96

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+ 31-70) 340-2040, Tx. 31 651 epo nl,
Fax (+ 31-70) 340-3016

Authorized officer

Cupido, M

INTERNATIONAL SEARCH REPORT

International Application No

PCT/GB 96/01341

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, vol. 90, no. 18, 15 September 1993, WASHINGTON US, pages 8324-8326, XP002016196 R.GRIFFITHS AND B.TIWARI: "The isolation of molecular genetic markers for the identification of sex" cited in the application see the sequence of primer ss2 see page 8324, right-hand column, paragraph 1 ---	3-5, 9-12, 14-18
X	WO,A,94 07907 (ZOOGEN,INCORPORATED) 14 April 1994 see table 2 ---	9-12, 15-18
X	DATABASE WPI Section Ch, Week 8849 Derwent Publications Ltd., London, GB; Class B04, AN 88-348691 XP002016199 & JP,A,63 258 580 (NICHIREI KK) , 26 October 1988 see abstract ---	9-12, 15-18
X	DATABASE WPI Section Ch, Week 8849 Derwent Publications Ltd., London, GB; Class B04, AN 88-348692 XP002016200 & JP,A,63 258 581 (NICHIREI KK) , 26 October 1988 see abstract ---	9-12, 15-18
X	PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, vol. 90, March 1993, WASHINGTON US, pages 2414-2418, XP002016197 V.DELMAS ET AL.: "A mammalian DNA-binding protein that contains a chromodomain and an SNF2/SWI2-like helicase domain" cited in the application see the whole document ---	19-21, 24,25
P,X	NATURE, vol. 375, 8 June 1995, LONDON GB, page 454 XP002016198 R.GRIFFITHS AND B.TIWARI: "Sex of the last wild Spix's macaw" see the whole document -----	3-5, 9-12, 14-18

INTERNATIONAL SEARCH REPORT

Information on patent family members

Inter-Application No

PCT/GB 96/01341

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO-A-9407907	14-04-94	CA-A- 2124220	14-04-94
		AU-B- 662564	07-09-95
		AU-A- 2696092	26-04-94
		EP-A- 0623139	09-11-94

THIS PAGE BLANK (USPTO)

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.

THIS PAGE BLANK (USPTO)